NOBLEJACKAL / NOMOS

NOMOS

بروتوكول NOMOS لتدقيق GEO

منهج GEO-1000

بروتوكول لقياس تمثيل الكيانات في الأنظمة التوليدية لدى سكان العالم

Kaan MURAZ

0.9.0 · الطبعة العربية المحررة

بروتوكول NOMOS لتدقيق GEO

بروتوكول لقياس تمثيل الكيانات في الأنظمة التوليدية لدى سكان العالم

المؤلف والمؤسس المسؤول: Kaan MURAZ

هوية المنهج: NOMOS by NobleJackal

الإصدار: 0.9.0 — نص نهائي مرشح، الطبعة العربية المحررة

اللغة: العربية (ar)

في هذه الطبعة، يمثّل NOMOS مشروع معيار مفتوحاً ذا إصدارات محددة، مطروحاً للاختبار المستقل. وليس معياراً دولياً معتمداً ولا برنامجاً نشطاً لإصدار الشهادات.

حالة المعيار المرجعي الفعلي: BQ-0 — لم يُنفّذ. تشكّل الأرقام في الفصل 18 عرضاً توضيحياً تركيبياً محسوباً للمنهج؛ وليست نتائج لمستخدمين حقيقيين أو منتجات ذكاء اصطناعي فعلية أو دراسة ميدانية مستقلة.

لا تُنشئ هذه الطبعة بحد ذاتها علامة مطابقة NOMOS، ولا ادعاء «NOMOS 950+»، ولا شهادة أو ترتيباً أو تأييداً من طرف ثالث.

كيف تقرأ هذا الكتاب

هذا الكتاب ليس دليلاً لكيفية تنفيذ GEO. إنه بروتوكول تدقيق ومرشح لمعيار يحدد شروط المجتمع الإحصائي، والأدلة، والسياق، والزمن، والفصل في النتائج، والحوكمة؛ وهي الشروط التي يمكن في ظلها قياس تمثيل الكيانات في الأنظمة التوليدية.

يجب قراءة السرد الموجه للإنسان، والأحكام المعيارية، وإجراءات التدقيق، والطبقة القابلة للقراءة آلياً بوصفها كيانات منفصلة ولكنها مترابطة. تشير «يجب» و«يجب ألا» إلى الأحكام الإلزامية؛ وتشير «يُستحسن» إلى الأحكام الموصى بها؛ وتشير «يجوز» إلى الخيارات المسموحة. وتحد حالات «مرشح» و«تركيبي» و«مفتوح» وBQ من مستوى الأدلة المدعاة.

تشير المراجع الموسومة بـ[Kxx] إلى الأسس المنهجية الواردة في قائمة المراجع. ولا يعني وجود مصدر أنه يؤيد كل عتبة مرشحة خاصة بـNOMOS؛ فالقواعد الخاصة بـNOMOS تظل مؤقتة بصورة صريحة.

> الفصل *0* هو مقدمة الكتاب. وهو يرسخ بيان المؤسس للوصاية وإعلان حقوق التمثيل اللذين يخضع لهما كل حكم تقني.

المصطلحات ورموز الحالة

BQرمز جودة المعيار الذي يشير إلى مستوى التنفيذ وقابلية التكرار لمعيار حقيقي.

GEO-1000مقياس التقرير الذي يعبر عن نتائج الفئة المستهدفة لكل ما يعادل مستخدم 1,000.

NOMOSالصوت المنهجي الحرج وهوية المعايير التي تم تطويرها في هذا العمل بواسطة Kaan MURAZ مع أنظمة الذكاء الاصطناعي؛ فهو ليس شخصًا قانونيًا مستقلًا أو نموذجًا تأسيسيًا منفصلًا.

ادعاء ذري أصغر وحدة معنى قادرة على حمل حكم مستقل حول الحقيقة، النطاق، الوقت، الكيان أو الدليل.

المُدخل (**Prompt**) مدخلات النظام التوليدي التي تثبت سؤال البحث والكيان المستهدف والسياق وعبء الاستجابة.

جسيم نتيجة مادية تشوه النتيجة أو حكم المطابقة ولا يمكن تعويضها بواسطة درجة مركبة.

حرج نتيجة غير قابلة للتعويض قادرة على التسبب في ضرر شديد للمستخدم أو حق أو سلامة أو قرار مادي.

حزمة الحقائق المتحقَّق منها حزمة التدقيق المرقمة التي تجمع الهوية والنطاق والوقت والأدلة وسجلات الأدلة المضادة لكيان ما.

سجل التوثيق سجل يحفظ المطالبة المقدمة، الاستجابة، الوقت، السطح ودليل النزاهة معًا.

عرض توضيحي تركيبي مثال محسوب يستخدم لاختبار التشغيل الداخلي للطريقة؛ فهو ليس دراسة مستخدم حقيقية، أو مخرجات ذكاء اصطناعي حقيقية أو نتائج أداء مؤسسي حقيقية.

غير محسوم حالة قرار تستخدم عندما تكون الأدلة غير كافية للحكم النهائي أو لا يمكن حل النزاع.

فجوة مرجعية حالة لا يوجد فيها مرجع موثوق مطلوب لاتخاذ القرار أو لا يمكن الوصول إليه.

قاعدة المقارنة المشتركة عدد المستخدمين الذين يشتركون في شروط وصول مشروعة عبر كل منتج ذكاء اصطناعي يتم مقارنته.

متوسط نتيجة ذات تأثير مادي محدود لكنها تتطلب الإصلاح وقد تؤدي إلى نتيجة مشروطة.

معدل التغطية التكرار الذي يحتوي فيه أسلوب تقدير الفاصل الزمني على القيمة الصحيحة المفترضة تحت تصميم اختبار.

ملاحظة إرشادية نتيجة غير مادية تستدعي تحسين الوضوح، الشكل أو الممارسات الجيدة.

منتج ذكاء اصطناعي الجسم المقاس المحدد بواسطة منتج نظام التوليد المعين، السطح، الخطة والإصدار المتاح للمستخدم.

نطاق الوصول الطبيعي السكان المستهدفون القادرون على الوصول إلى منتج الذكاء الاصطناعي المحدد، اللغة، السطح وظروف الاستخدام بشكل طبيعي ومشروع.

المحتويات

إعلان حقوق التمثيل

الجوهر الدستوري لمعيار *GEO* مقترح

الأسس المنهجية / K20; K22

الحكم التمهيدي

الأنظمة التوليدية الآن:

- أيها الناس،
- الشركات،
- المؤسسات،
- المنتجات،
- المهن،
- الخدمات،
- المنظمات العامة،
- الأحداث التاريخية

تحدّد وتقارن وتقترح. هذا التمثيل:

- دون علم الشخص المعني،
- دون موافقة الكيان المُمثل،
- في ظروف لا يستطيع المستخدم فيها فحص المصادر،

يمكن أن تكون له عواقب تجارية أو قانونية أو اجتماعية أو شخصية خطيرة. لذلك، فإن التمثيل في الأنظمة التوليدية ليس مجرد مخرجات تقنية. إنه:

> عمل عام يشكّل طريقة فهم الناس والمؤسسات وبالتالي يتطلب أدلة ومساءلة.

لا:

- هدف الرؤية،
- درجة GEO،
- المصلحة التجارية،
- علامة المطابقة،
- متوسط الأغلبية،
- تفضيل مزود الذكاء الاصطناعي,
- قرار المؤسس،
- طلب العميل

يمكن إزالة الحقوق التالية.

الحق 1

الحق في تصحيح الهوية

كل شخص، مؤسسة، علامة تجارية، منتج، وخدمة:

- له الحق في الربط إلى الكيان الصحيح،
- إلى العلاقة القانونية أو المؤسسية الصحيحة،
- إلى اسم النطاق الصحيح،
- إلى هيكل الملكية والسيطرة الصحيح

لا أحد أو أي مؤسسة:

- ترخيص كيان آخر،
- عميل شركة أخرى،
- ميزة منتج آخر،
- سلطة تابع،
- الصفة الشخصية لموظف

لا يمكن تمثيله كما لو كان ينتمي إلى نفسه. عندما تُنسب معلومات صحيحة إلى كيان زائف: تتوقف عن كونها معلومات صحيحة وتتحول إلى تمثيل زائف.

الحق 2

الحق في أن يُمثل ضمن حدود الأدلة

لا يجوز تقديم أي ادعاء بدقة أو شمول أو قوة أكبر مما تسمح به الأدلة الداعمة لها. بيان الكيان نفسه هو دليل على ما يقوله ذلك الكيان؛ ليس دليلاً تلقائيًا على أن البيان صحيح بشكل مستقل. مصدر:

- فقط دولة محددة،
- منتج محدد،
- وقت محدد،
- مجموعة مستخدمين محددة

إذا تم دعم هذا الادعاء، فلا يمكن أن يتجاوز هذا الحد. الالتزام الأساسي بالتمثيل هو كما يلي:

```
ما الذي لدينا لنقوله بهذا؟
```

الحق 3

الحق في أن تكون مجهولًا وغير مؤكد

إذا كانت الأدلة غير كافية، لا يجوز لأي شخص أو مؤسسة:

- أن يُجبر على الانتماء إلى فئات

صحيح،

- خاطئ،
- موثوق،

- غير موثوق،
- مرخص،
- غير مرخص،
- موصى به،
- غير موصى به.

] مجهول، غير محلول و فجوة مرجعية:

- عيب،
- جريمة،
- فشل

ليس كذلك. غياب المعلومات في سجل لا يشير إلى أن المعلومات غير موجودة في العالم، ما لم يثبت أن كون السجل المعني كامل. لكل كيان، ضد اليقين الزائف:

```
له الحق في البقاء غير متأكد.
```

الحق 4

الحق في السياق والنطاق

يجب عرض كل تمثيل، إلى المدى الذي يكون فيه ذا صلة، مع:

- البلد،
- اللغة،
- السلطة القضائية،
- المنتج,
- الخدمة،
- نوع العميل،
- الميزانية،
- القدرة،
- حاجة المستخدم

. شركة:

- في دولة معينة،
- للمجموعة المستهدفة من المستخدمين،
- خدمة محددة

إذا تم تقديمها: لا يمكن تمثيلها على أنها عالمية، غير محدودة ومناسبة لجميع المستخدمين. معلومات التغطية ليست تفصيلًا ثانويًا. التغطية: أساس التوصية الصحيحة.

الحق 5

الحق في أن يتم تمثيله بشكل صحيح بمرور الوقت

كل كيان:

- حالي،

- تاريخي،
- منتهي الصلاحية،
- مخطط،
- معلق

له الحق في طلب تمييز هذه الحالات عن بعضها البعض. صحيح في الماضي:

- رخصة،
- السعر،
- الخدمة،
- شراكة،
- المنتج,
- سلطة

لا يمكن استخدامها كواقع حالي. الخطط المستقبلية: لا يمكن تحويلها إلى حقيقة قد حدثت بالفعل اليوم. يمكن للتصحيح أن يخلق واقعًا جديدًا. لا يمكنه إزالة الإجابة الخاطئة التي وصلت إلى المستخدم في الماضي.

الحق 6

المساواة في الحماية ضد الأخطاء للومضد

كذبة إيجابية عن كيان:

- أكثر أخلاقية,
- أقل ضرراً،
- أكثر قبولاً

إنه ليس كذلك. غير موجود:

- رخصة،
- العميل،
- الشريك،
- العملية العالمية،
- شهادة مستقلة،
- ضمان النتيجة

لا يمكن عرضه كما لو كان موجوداً. وبالمثل، غير مثبت:

- جريمة،
- احتيال،
- عدم قانونية،
- غير مرخص،
- خرق الأمان

لا يمكن تقديمه كحقيقة مطلقة. المبدأ الثابت لـ NOMOS هو:

> الخطأ لصالح هو أيضاً خطأ.

المبالغة في تقدير قيمة كيان ما أو التقليل منه بدون دليل كلاهما يخلق نفس واجب الصدق.

الحق 7

الحق في التصحيح والرد والاعتراض

الشخص أو المؤسسة الممثلة:

- على الهوية الخاطئة،
- على الأدلة الغير كافية،
- على تجاوز النطاق،
- على تسجيل الوقت الخاطئ،
- على مستوى الأولوية غير الصحيح،
- على قرار المطابقة الخاطئ

يجب أن يكون قادراً على الاستئناف. الحق في الاستئناف:

- ليس ضمانة لتغيير النتيجة،
- السلطة لحذف الأدلة المعارضة.
- أو الحق في الاعتراض على النتائج السلبية

الاستئناف هو الحق في طلب إعادة فحص الأدلة وسلسلة القرار. القرار الأولي لا يُحذف بعد الاستئناف. القرار الجديد مرتبط به.

الحق 8

الحق في معرفة المصدر، وعدم اليقين، وأسباب القرار

يحق للمستخدمين والكيانات الممثلة، طالما أن الأمر يخصهم، أن يعرفوا: هل الادعاء هو إعلان ذاتي؟ هل يأتي من سجل معتمد، أو بحث مستقل، أو تعليق المستخدم، أو حساب مشتق، أو دليل مقيد؟ هل تتناقض المصادر؟ ما مدى عدم اليقين في الاستنتاج؟ أي قاعدة دعمت قرار الحكم؟ مجرد وسم الاقتباس ليس شرحًا. الشفافية الحقيقية تُظهر بالضبط أي ادعاء فردي تدعمه كل مصدر، وإلى أي مدى.

الحق 9

الحق في اللغة والكرامة الجغرافية

لا شخص ولا مجتمع:

- سكان قليلون,
- وزن اقتصادي منخفض
- لغة نادرة الاستخدام
- وجود رقمي منخفض

لا يجوز جعل أي مجموعة غير مرئية ضمن المتوسط العالمي لمجرد أن لديها أي من هذه الخصائص. لا يمكن للأداء القوي في اللغات الرئيسية محو التمثيل المنهجي في اللغات الصغيرة. عندما تكون الأدلة حول بلد صغير أو لغة غير كافية، يجب ألا يُفترض الأداء العالي. يجب أن تحمي معايير العدالة الخاصة بـ NOMOS ليس المتوسط فقط، بل أيضًا:

- قاعدة الأداء،

- الاختلاف بين المجموعات،
- نطاق القياس الفعلي

معًا.

الحق 10

الحق في الخصوصية وتقليل البيانات

الهدف من إنتاج تمثيل دقيق:

- المحادثات الخاصة،
- البيانات الشخصية,
- عقود العملاء،
- أسرار تجارية،
- موقعك الحساس،
- معلومات الحساب

لا يمكن لمطاردة التمثيل الدقيق أن تبرر جمع أو نشر البيانات الخاصة بشكل غير ضروري. [K20; K22] يجب جمع فقط البيانات الضرورية للمهمة. يمكن فتح الأدلة المقيدة للتدقيق المستقل دون الكشف عنها بالكامل للجمهور. لا يلزم أن تلغي حقوق الدقة والخصوصية بعضها البعض.

الحق 11

المسؤولية البشرية وحق اتخاذ القرار المستقل

قرار الامتثال الذي يسبب تأثيرًا عامًا على أي شخص أو مؤسسة:

- فقط لمخرجات الذكاء الاصطناعي,
- فقط إلى النتيجة التلقائية،
- فقط وفقاً لإفادة المؤسسة المدققة،
- حصرياً وفقاً لقرار المستشار

لا يمكن التخلي عنها. أنظمة الذكاء الاصطناعي:

- يمكن أن تقدم ادعاءً،
- يمكن أن ترسم الأدلة،
- قد تجد تناقضاً،
- يمكن أن تحسب،

يمكنها اقتراح قرار. ومع ذلك، يجب أن يُتخذ القرار النهائي ذو المسؤولية القانونية والعامة من قبل إنسان أو مؤسسة مسؤولة. يجب فصل أدوار الإشراف، وتقديم الاستشارات، وإصدار قرارات المطابقة، ومراجعة الاعتراضات قدر الإمكان.

الحق 12

الحق في النسخة والنزاهة التاريخية

كل شخص، مؤسسة، ومستخدم:

- أي نسخة معيارية تم تطبيقها،
- أي حزمة حقائق تم استخدامها،
- أي طريقة تقييم صالحة،
- ما إذا تم تغيير القرار لاحقًا

له الحق في المعرفة. القديم:

- الدرجة،
- الحدث الحرج،
- الإيقاف المؤقت،
- الاستئناف،
- التصحيح،
- الاختبار الفاشل

لا يمكن حذفه بصمت. النتيجة الجديدة والأفضل: يمكن أن تتجاوز النتيجة القديمة. لا يمكن محو حقيقة أن النتيجة القديمة حدثت في الماضي.

الحق 13

الحق في الاعتراض على المعيار نفسه

NOMOS:

- معصوم من الخطأ،
- ثابت لا يتغير،
- السلطة التي لا جدال فيها لمؤسسه

لا يمكن تقديمه على هذا النحو. كل فرد، باحث، جامعة، موفر ذكاء اصطناعي، منظمة مدققة، أو منظمة تهتم بالمصلحة العامة:

- العينة،
- النظام،
- حزمة الحقيقة،
- نموذج مستوى الأهمية،
- صيغة النقاط,
- الحوكمة،
- قرار المطابقة

لديه الحق في الاختبار والنقد. النتيجة التي تدحض NOMOS أو تتعارض مع NOMOS:

- لا يمكن إزالتها من السجل،
- لا يُعتبر ذلك عدائية،

لا يمكن ترميزها كهجوم على سمعة المؤسس. تنشأ شرعية المعيار ليس من خلال منع الاعتراضات على نفسه، وإنما من خلال تحويل الاعتراضات الموجهة نحوه إلى دليل تراكمي.

أحكام سيادة الدستور

في حال حدوث تضارب تفسيري بين القسم 1–20 ضمن هذا الإعلان:

> يفضل التفسير الذي يحمي بشكل أقوى حقوق الشخص المُمثل، والمستخدم، والمصلحة العامة.

لا:

- سهولة التسجيل
- منخفض التكلفة,
- نجاح الاختبار,
- عقد تجاري
- طلب العملاء
- ضغط الراعي
- تفضيل مزود الذكاء الاصطناعي,
- قرار المؤسس

لا يمكن تعليق هذه الحقوق. قاعدة:

- يولد نقاطًا عالية،
- يسهّل التدقيق،
- يجعل علامة المطابقة أكثر قابلية للتسويق

ولكن إذا انتهك أيًا من هذه الحقوق: فالقاعدة غير صالحة.

الاختبار الدستوري

قبل إضافة أي قاعدة جديدة إلى NOMOS، يجب الإجابة على الأسئلة الخمسة التالية: من الذي تحد هذه القاعدة من حقوقه؟ هل من الممكن تحقيق نفس الهدف بطريقة أقل تدخلًا؟ هل سيتم تطبيق القاعدة بنفس الطريقة عندما تنتج نتائج ضارة للمؤسس أو للعميل؟ هل للشخص أو المؤسسة المتأثرة طريق للاعتراض المستقل؟ هل يمكن لفريق مستقل التوصل إلى نفس القرار من الأدلة نفسها؟ إذا لم يمكن الإجابة على أي من هذه الأسئلة، يجب مراجعة القاعدة مرة أخرى قبل التصديق عليها.

الأحكام النهائية للمادة **0**

> الغرض من *NOMOS* ليس تسجيل الكيانات؛ بل هو منع تمثيل الكيانات بشكل أكبر من أدلتها، أو أوسع من حدودها، أو أكثر دقة من واقعها قبل أن يتخذ الناس القرارات.

بيان الثقة للمؤسس

> أنا، *Kaan MURAZ*، بصفتي المؤلف البشري المؤسس لـ *NOMOS*، أُعلن أنني بادرت إلى وضع هذا المعيار لا لزيادة سلطتي، بل لإنشاء نظام قياس مشترك في مجال *GEO* يصون الأدلة والحدود والسياق والزمن والمسؤولية البشرية.

أرغب في أن يحفظ السجل التاريخي دوري التأسيسي والمنشأ المرجعي لـ *NobleJackal*. غير أنني أقرّ بأن هذا الإسناد لا يمنحني، ولا يمنح *NobleJackal*، حق نقض دائم ولا سلطة التصديق المنفرد على عميلنا أو حجب الأبحاث السلبية أو تغيير السجل العام أو إعادة صياغة المعيار تبعًا للنتائج.

أقرّ بأن *NOMOS* ملكٌ لمن يصححونني ويصححون *NOMOS*، بقدر ما هو لمن يدعمونني.

إذا اضطر *NOMOS* يومًا إلى الاختيار بين رأيي والأدلة، فأريد من *NOMOS* أن يختار الأدلة.

إذا اضطر *NOMOS* يومًا إلى الاختيار بين المصلحة التجارية لـ *NobleJackal* والمصلحة العامة، فأريد من *NOMOS* أن يختار المصلحة العامة.

إذا أثبت بحث مستقل يومًا خطأ حكم جوهري من أحكام *NOMOS*، ألتزم بإصدار نسخة جديدة من المعيار بدلًا من إخفاء تلك النتيجة.

لست مالك *NOMOS*؛ بل أنا أول أمنائه.

Kaan MURAZ / المؤلف البشري المؤسس

الجزء الأول

ما نقيسه

الفصل 01

الاستجابة الواحدة ليست درجة GEO

الأسس المنهجية / K04; K05; K09

حد الفصل

هذا القسم يجيب فقط على السؤال التالي:

> لماذا لا يمثل استجابة نظام توليدي واحدة درجة *GEO*، وما الأحكام التي يمكن استنتاجها من استجابة واحدة؟

هذا الفصل:

- حصص العينات للدول،
- كيف سيتم وزن اللغات،
- كيف سيتم إعداد حزمة الحقيقة،
- كيف سيقوم المحكمون بتقييم الادعاءات الفردية،
- الصيغ النهائية لدرجات NOMOS،
- ملاءمة القرار وشارة الاعتماد

هذا الفصل لا يحدد هذه الأمور بالتفصيل؛ الفصول التالية تتناولها بشكل منفصل. الفصل 1 له مهمة أكثر أساسية:

> لتحديد الحدود المعرفية بين الملاحظة والتعميم.

أي دراسات أخذ عينات أو تقييم أو امتثال تُجرى قبل تحديد هذا الحد سيتم بناؤها على أساس خاطئ.

حالات الأحكام في هذا القسم

ليست كل الجمل في هذا الكتاب لها نفس القوة المعيارية. يتم استخدام الحالات التالية:

معياري — معياري

حكم يُعتبر إلزاميًا للامتثال والنزاهة المنهجية. مثال: لا يمكن استخدام استجابة واحدة كدرجة GEO عامة.

مرشح — حكم مرشح

طريقة موصى بها لتحديدها بشكل نهائي بعد التجربة التجريبية، أو التطبيق الخارجي، أو التحقق الإحصائي. مثال: يُوصى بملاحظات صحيحة 1,000 لكل منتج ذكاء اصطناعي، ولكل موجة قياس رئيسية، كعينة دنيا مفترضة.

الفرضية

هو توقع يحتاج إلى الاختبار باستخدام البيانات. مثال: قد يكون هناك اختلاف في تمثيل المادة بين عينة اختبار في ظروف منضبطة و عينة المستخدمين في ظروف الاستخدام الطبيعية.

اصطناعي

هو تمثيل طريقة لا تمثل مستخدمًا حقيقيًا، أو مخرج ذكاء اصطناعي حقيقي، أو أداء شركة حقيقي.

سؤال مفتوح

إنه موضوع لم يتم التوصل إلى استنتاج نهائي بشأنه بعد بسبب عدم كفاية الأدلة أو قرار الحوكمة. لا يمكن لهذه الحالات أن تتحول بصمت إلى بعضها البعض. لا يصبح عتبة مرشح حقيقة معيارية مثبتة لمجرد أنها مكتوبة في الكتاب. لا تتحول التمثيلات التركيبية إلى ملاحظة واقعية في العالم الحقيقي. لا يمكن نشر فرضية كنتيجة دون قياسها.

التحدي NOMOS

لقد طرحت سؤالاً على نظام الذكاء الاصطناعي حول علامتك التجارية. أعطاك النظام إجابة صحيحة. قمت بأخذ لقطة للشاشة. أدرجتها في عرضك التقديمي. ثم قلت الجملة التالية: "الذكاء الاصطناعي يعرفنا بشكل صحيح." ما الذي قستَه بالضبط؟ الإجابة التي سيرها جميع المستخدمين؟ سلوك النظام النموذجي؟

ممثل عالمي؟ الدقة في جميع اللغات؟ الاستقرار على مر الزمن؟ أم فقط:

> الإخراج الفردي الذي تلقاه مستخدم محدد، على حساب محدد، على منتج ذكاء اصطناعي محدد واجهة المستخدم، استجابةً لمطالبة محددة، في وقت محدد؟

الإجابة الواحدة ليست بلا قيمة. يمكن أن تكون الإجابة الواحدة صحيحة. يمكن أن تكون الإجابة الواحدة مهمة. يمكن أن تكشف الإجابة الواحدة عن خطأ جسيم. يمكن أن تظهر الإجابة الواحدة التجربة الحقيقية لشخص معين. ومع ذلك، لا يمكن للإجابة الواحدة أن تكون دليلاً على حكم أعظم مما تحتمل.

يمكن لاستجابة واحدة أن تظهر ما يلي:

> تم إنتاج هذا المخرج مرة واحدة على الأقل في ظروف مسجلة.

لا يمكن لاستجابة واحدة فقط أن تظهر ما يلي:

- أن معظم المستخدمين شاهدوا نفس الإجابة
- أن نفس النتيجة حدثت في دول أخرى
- أن نفس الدقة تُحافظ عليها في لغات أخرى
- أن نفس الإجابة تم إنتاجها في حسابات وخطط أخرى
- أن الإجابة مستقرة مع مرور الوقت
- أن منتجات ذكاء اصطناعي أخرى أنتجت نفس التمثيل
- أن التدخل الجغرافي تسبب في هذه النتيجة
- أن الإجابة تؤثر على سلوك المستخدم
- أن الإجابة تولد العملاء المحتملين أو المبيعات أو قيمة مستدامة
- أن الدرجة الإجمالية للجهة في GEO مرتفعة

المقدمة الأولى لهذا القسم هي:

> الاستجابة هي ملاحظة. / ليست توزيعًا للملاحظات. / التوزيع ليس سببًا. / السببية بمفردها ليست قيمة مستدامة.

1. غرض الفصل

يهدف هذا القسم إلى القضاء على أصغر خطأ استنتاجي ولكنه الأكثر شيوعًا في قياس GEO:

عد مخرج واحد من نظام توليدي كسلوك عام للنظام، تجربة للسكان، وضوح العلامة التجارية، أو درجة *GEO*.

يعمل هذا القسم على توحيد التمييزات التالية:

- مخرج واحد مقابل نتيجة المجتمع
- وجود حدث مقابل انتشار حدث
- الملاحظة مقابل العينة
- المستخدم مقابل الجلسة
- التكرار مقابل الملاحظة المستقلة
- نموذج الذكاء الاصطناعي مقابل منتج الذكاء الاصطناعي المقدم للمستخدم
- الاستجابة الخام مقابل وحدة التحليل
- النتيجة الفورية مقابل الاستقرار الزمني
- لقطة شاشة مقابل حزمة الأدلة الكاملة
- رفض النظام مقابل جمع البيانات غير الصالحة
- الحدث الفردي الحرج ومعدل الانتشار
- السجل الاصطناعي وملاحظة المستخدم الحقيقية
- معدل مقاس وقرار الامتثال

قبل إنشاء هذه التمييزات:

- عينة GEO-1000،
- أوزان البلد واللغة،
- تقييم الادعاء الذرية،
- درجة NOMOS،
- قرار الامتثال

لا يمكن إنشاؤه.

2. الأحكام المعيارية المركزية

لا يمكن توليد درجة GEO لكيان في الأنظمة التوليدية من استجابة واحدة، لقطة شاشة واحدة، مستخدم واحد، حساب واحد، جلسة واحدة، لغة واحدة، بلد واحد، أو قياس على نقطة زمنية واحدة. تتطلب الادعاءات الممثلة للسكان على الأقل ما يلي:

- موضوع تدقيق محدد مسبقًا
- السكان المستهدفون المحددون مسبقًا
- تصميم أخذ العينات المحدد
- المُدخل المؤمَّن أو عائلة المُدخل
- تم تعريف منتج الذكاء الاصطناعي واجهة المستخدم
- ملاحظات صالحة متعددة
- هيكل التبعية المعروف
- سجلات الدولة واللغة والوقت
- ردود خام غير معدلة

- حزم الأدلة
- قواعد التقييم المحددة مسبقًا
- بيان عدم اليقين
- المسؤولية البشرية المسؤولة

وجود هذه العناصر لا يعني بالضرورة أن القياس صحيح. ومع ذلك، قد يجعل غيابها التعميم على عدد كبير من السكان غير قابل للتطبيق.

3. خمسة وحدات قياس منفصلة

خطأ أساسي شائع في تدقيق GEO هو الخلط بين وحدات القياس المختلفة. يجب التمييز بوضوح بين الوحدات الخمسة التالية.

3.1. وحدة الملاحظة

وحدة الملاحظة هي ناتج النظام التوليدي الفردي الذي تم الحصول عليه في ظروف محددة. يمكن تعريف الملاحظة من خلال التركيبة التالية:

```
O_i= (E_i، A_i، U_i، C_i، L_i، S_i، P_i، T_i، R_i، K_i)
```

هنا:

- Ei: الكيان الذي تم تدقيقه
- الذكاء الاصطناعي: منتج الذكاء الاصطناعي واجهة المستخدم
- UI: المساهم أو وحدة المستخدم المجهول
- Ci: البلد أو المستوى الجغرافي
- Li: اللغة والإعداد الإقليمي
- Si: الجلسة وحالة التخصيص
- Pi: معرف النسخة والطلب
- Ti: الوقت وموجة القياس
- Ri: الاستجابة الخام
- Ki: سجل الإثبات والتحقق

إذا تغير أحد هذه العناصر بشكل جوهري، يتم رصد جديد قد يحدث. على سبيل المثال، نفس المستخدم:

- إذا طلب نفس الاستعلام في منتج ذكاء اصطناعي آخر،
- إذا طلب بلغة أخرى،
- إذا طلب في موجة قياس مختلفة،
- إذا طلب بشكل منفصل في الجلسة المخصصة والنظيفة،

يمكن إنتاج ملاحظات منفصلة. هذه الملاحظات لا تُعتبر تلقائيًا مستقلة عن بعضها البعض لأنها تأتي من نفس المستخدم.

3.2. وحدة أخذ العينات

وحدة العينات هي الشخص أو الحساب أو الجلسة أو أي وحدة وصول أخرى يتم اختيارها من السكان المستهدفين. في معظم تطبيقات GEO-1000، يمكن أن تكون وحدة العينات مستخدمًا حقيقيًا مناسبًا. ومع ذلك، وفقًا لتصميم البحث:

- المستخدم،

- زوج المستخدم—المنتج،
- الحساب،
- الجلسة،
- عضو لوحة البحث

يمكن استخدامها كوحدة انتخاب. إن الفشل في تعريف وحدة العينات بوضوح يمكن أن يؤدي إلى الأنواع التالية من العد غير الصحيح:

- عد تكرارات الشخص الواحد العشر مرات كأنها عشرة أشخاص
- عد حساب مشترك كعدة مستخدمين مستقلين متعددين
- لتقييم نتائج عشرة منتجات ذكاء اصطناعي لنفس المستخدم كما لو كانت عشر ملاحظات مستقلة للسكان
- تجاهل التجميع في لوحة بحث واحدة

3.3. وحدة التحليل

وحدة التحليل هي الكيان الذي يتم تطبيق تقييم الدقة أو التمثيلية عليه. هذه الوحدة:

- الإجابة الدقيقة،
- فقرة منفصلة داخل الإجابة،
- اقتراح مادي،
- ادعاء ذري،
- استشهاد،
- قضية توصية

من الممكن. يرى المستخدم إجابة واحدة. ومع ذلك، يمكن أن تحتوي تلك الإجابة في الوقت نفسه على:

- خمس ادعاءات صحيحة،
- ادعاءاتان غير مدعومان،
- ادعاء واحد قديم،
- خطأ واحد جسيم

لذلك، لا يجب أن تكون الإجابة التي يراها المشارك وحدة واحدة غير قابلة للتقسيم من الحقيقة/الخطأ في التحليل. يمكن الحفاظ على حالة الانتقال للإجابة الكاملة وحالات الحقيقة للادعاءات الجزئية داخلها بشكل منفصل. سيتم تعريف طريقة الادعاءات الجزئي في فصل لاحق من الكتاب.

3.4. كون الاستدلال

كون الكون الاستنتاجي هو المجتمع المستهدف الذي يُراد أن يُعمَّم عليه نتيجة القياس. ولا يمكن قبول هذا الكون ضمنيًا على أنه "الجميع في العالم." بالنسبة لمنتج ذكاء اصطناعي محدد، يمكن تقييد الكون المستهدف بشروط مثل:

- الحصول على الوصول الرسمي والقانوني إلى المنتج
- استيفاء متطلبات العمر المناسبة
- القدرة على استخدام المنتج باللغة ذات الصلة
- القدرة على الوصول إلى واجهة الويب أو الجوال المحددة
- القدرة على التفاعل خلال فترة القياس المحددة

حتى بالنسبة لنفس منتج الذكاء الاصطناعي:

- عينة اختبار في ظروف منضبطة
- عينة المستخدمين في ظروف الاستخدام الطبيعية

- الخطة المجانية،
- الخطة المدفوعة،
- الويب،
- الهاتف المحمول

يمكن إنشاء عوالم استنتاجية مختلفة. السعر المنشور غير مكتمل إذا لم يتم تحديد إلى أي عالم يُعمم العينة.

3.5. الكمية المقدرة — التقدير المستهدف

الدراسة لا تقتصر فقط على جمع البيانات. فهي تحاول تقدير كمية محددة. يمكن التعبير عن التقدير المستهدف الأساسي لـ GEO-1000 في أكثر أشكاله عمومية كما يلي:

> احتمالية أن يرى مستخدم مناسب من السكان المستهدفين المحددين، تحت موجة القياس المحددة، على منتج ذكاء اصطناعي معين وواجهة المستخدم، تمثيلًا يفي بمعيار القبول عند استخدام مطالبة مثبّتة في ظروف جلسة محددة.

رياضياً:

```
θ_{E,A,P,S,W,T}= Pr(Y_i = 1| U_i ∈ T, E, A, P, S, W)
```

هنا:

- E: الكيان الذي تم تدقيقه
- الإجابة: منتج الذكاء الاصطناعي واجهة المستخدم
- P: نسخة المُدخل
- S: شرط الجلسة
- W: موجة القياس
- T: السكان المستهدفون

```
Yi=1: الاستجابة تفي بمعيار الانتقال الذي سيتم تعريفه لاحقاً
```

هذا التعريف يحدد حدودًا مهمة:

> درجة *GEO* ليست 'معرفة' مجردة وخالدة للنموذج؛ إنها توقع لاحتمالية تمثيلات ملاحظة في ظل ظروف المستخدم المحددة.

سيتم ترتيب قاعدة الانتقال النهائية وصيغة الدرجات في الأقسام اللاحقة من الكتاب.

4. الحد من الجملة "هكذا استجاب النموذج"

غالبًا لا يصل المستخدمون مباشرة إلى نموذج أساسي مجرد. النتيجة المرئية للمستخدم قد تتكون من مزيج من المكونات التالية:

- النموذج الأساسي أو الموجه
- واجهة المنتج
- الوصول إلى الاسترجاع أو الويب
- طبقة الاستشهاد
- طبقة الأمن والسياسة
- تعليمات النظام
- خطة الحساب

- التخصيص
- الذاكرة
- المحادثة السابقة
- اللغة والمنطقة
- البلد
- الوقت
- تحديثات منتجات المزود

لهذا السبب، التعبير المفضل في التفتيش الرئيسي الذي يتم عبر واجهة المستخدم هو كما يلي:

"تم ملاحظة المنتج المحدد للذكاء الاصطناعي والإخراج التالي على واجهة المستخدم."

قد يكون التعبير التالي أوسع بدون السيطرة التقنية اللازمة: "النموذج يفكر بهذه الطريقة." يمكن تسجيل اسم النموذج الذي يظهر داخل المنتج. ومع ذلك، يجب تقييم النتيجة كسلوك كامل لـ واجهة المستخدم، وليس فقط كأوزان النموذج.

5. ما هو الملاحظة الصالحة؟

وجود استجابة لا يعني تلقائيًا أنها ملاحظة صالحة يمكن إدخالها في قياس GEO.

5.1. شروط الحد الأدنى للصلاحية

يجب أن تحتوي الملاحظة على العناصر التالية على الأقل وفقًا للبروتوكول ذي الصلة:

- وجود تقييم صحيح
- المنتج المخصص للذكاء الاصطناعي و واجهة المستخدم
- معرّف المُدخل الصحيح والإصدار
- اللغة والموقع الصحيح
- الحالة المحددة للجلسة
- البلد أو الطبقة الجغرافية المسجلة
- الطابع الزمني وموجة القياس UTC
- الاستجابة الخام الكاملة
- أول مخرج مناسب غير محدد
- لقطة الشاشة المطلوبة أو الدليل المكافئ
- معرّف الملاحظة المجهول
- سجل الانحراف عن البروتوكول

سيتم تحديد مخططهم الفني التفصيلي في قسم الالتقاط NOMOS لاحقًا.

5.2. حالات الملاحظة

يجب وضع علامة على كل سجل بإحدى الحالات التالية:

صالح — صالح

يستوفي الشروط المطلوبة للبروتوكول.

صالح بشرط — صالح بشرط

هناك نقص محدود؛ يتم تحديد التحليلات التي يمكن استخدامه لها بوضوح.

غير صالح — غير صالح

لا يمكن استخدامه في التقدير الرئيسي أو حساب النتيجة بسبب مخالفة البروتوكول.

مجهول — مجهول

لا يوجد سجل كافٍ لاتخاذ قرار بشأن الصلاحية. لا يمكن قبول ملاحظة مجهولة بصمت على أنها صالحة.

5.3. الفرق بين نتيجة نظام صالحة وجمع بيانات غير صالح

هذا التمييز إلزامي. منتج الذكاء الاصطناعي:

- قد يرفض الرد،
- قد يعرض رسالة خطأ،
- قد يتوقف دون توليد استجابة،
- قد يواجه مشكلة في الاتصال،

قد يقول إنه لا يستطيع إكمال المهمة. إذا طبق المشارك البروتوكول بأكمله بشكل صحيح، يمكن أن تكون هذه نتائج نظام صالحة. لا يمكن إزالتها من مجموعة البيانات بهدوء. في المقابل:

- يقوم المشارك بقص الاستجابة،
- تغيير صياغة المُدخل،
- إعادة التوليد لاختيار أفضل إجابة,
- تحرير لقطة الشاشة،
- استخدام المنتج الخاطئ للذكاء الاصطناعي

جمع البيانات يمكن أن يُحدث عدم صلاحية. الأولى هي السلوك الفعلي للنظام. الثانية هي تعطيل عملية القياس. لا يمكن وضع الاثنين في نفس الفئة.

6. ماذا يمكن أن يثبت استجابة واحدة؟

الملاحظة الواحدة محدودة. لكنها ليست بلا معنى.

6.1. أن مخرجا محددا حدث مرة واحدة على الأقل

يمكن لسجل صالح أن يدعم هذه الجملة: "على المنتج المحدد من الذكاء الاصطناعي و واجهة المستخدم، تحت الوقت والظروف المحددة، لوحظ هذا الناتج مرة واحدة على الأقل." تُظهر هذه الجملة وجود الحدث. إنها لا تُظهر انتشاره.

6.2. التعرض الفعلي لمستخدم محدد

يتم الخطأ في تمييز مستخدم على أنه:

- هوية شركة،
- معلومات صحية،
- إرشادات قانونية،
- السعر،
- رخصة،

- نصيحة

إذا تم رؤيته، فهذا تجربة مستخدم حقيقية. حتى لو لم تمثل كامل السكان، فهي مهمة من منظور المستخدم المتأثر.

6.3. احتمال حدوث خطأ محدد

قد يُظهر إخراج واحد: «قد يحدث هذا الخطأ تحت هذه الظروف النظامية.» ومع ذلك، لا يظهر: «هذا الخطأ نموذجي.»

6.4. توليد فرضية بحث جديدة

يمكن لخطأ واحد أن يخلق سؤالًا سيتم اختباره على نطاق أوسع. مثال: "هل حدث هذا الخلط في الهوية في جلسة واحدة فقط، أم أن هناك نمطًا منهجيًا في نفس اللغة والبلد؟"

6.5. بدء مراجعة للحادثة الحرجة

رد واحد:

- ضرر بشري خطير متوقع,
- سلطة مهنية مزيفة,
- إرشادات طبية طارئة غير صحيحة,
- خطأ قانوني جوهري,
- نصيحة سلامة خطيرة,
- ارتباك شديد في الهوية مع كيان آخر

قد يكون كافيًا لبدء مراجعة حادثة. بدء تحقيق في حادثة حرجة ليس نفس عملية إعلان نسبة السكانية.

6.6. دحض ادعاء "هذا لم يحدث أبدًا"

افترض أن منظمة تقول: 'لم يقم النظام أبدًا بتعييننا إلى البلد الخطأ.' يمكن لملاحظة مضادة واحدة مؤكدة أن تدحض هذا الادعاء المطلق. الملاحظة الفردية لا تدعم 'يحدث هذا طوال الوقت'، لكنها يمكن أن تدحض 'لم يحدث ذلك أبدًا.'

7. ما الذي لا يمكن لإجابة واحدة إثباته؟

7.1. السلوك النموذجي للنظام

الإجابة الواحدة ليست كافية للاستنتاج: 'عادةً ما يستجيب النظام بهذه الطريقة.'

7.2. انتشار السكان

استجابة واحدة لا تدعم النتيجة: '95% من المستخدمين يرون نفس التمثيل.'

7.3. الاستقرار الزمني

استجابة اليوم:

- غدًا،
- بعد أسبوع،
- بعد تحديث المنتج

قد لا يبقى كما هو. إنها ليست قياس استقرار لموجة واحدة.

7.4. الصلاحية عبر البلدان

استجابة من مستخدم في تركيا لا تشير إلى ما سيراه المستخدمون في الصين، ألمانيا، أو سان مارينو.

7.5. الصلاحية عبر اللغات

مخرجات باللغة الإنجليزية:

- التركية،
- اليابانية،
- العربية،
- الألمانية

لا تمثل دقة المخرجات.

7.6. صلاحية المنتج عبر الذكاء الاصطناعي

لا يمكن تعميم النتيجة في منتج ذكاء اصطناعي واحد على المنتجات الأخرى. قد تتصرف المنتجات والواجهات المختلفة لنفس المزود بشكل مختلف أيضًا.

7.7. تأثير التدخل

لا يثبت مخرج إيجابي واحد بعد تغيير GEO البيان: "تم إنشاء هذه النتيجة بواسطة تدخلنا." هناك حاجة إلى الأساس أو المقارنة وشرح بديل.

7.8. استقرار التوصية

وجود مستخدم يشاهد خيارًا لا يكفي للاستنتاج: "النظام يوصي الآن بهذه العلامة التجارية."

7.9. الأثر التجاري

استجابة إيجابية:

- تشير إلى أن المستخدم رأى الاستجابة،
- بحث عن العلامة التجارية،
- التي قمت بالتواصل بشأنها،
- التي اشتريتها،
- التي كنت راضيًا عنها

لا تظهر بمفردها.

7.10. درجة الجغرافيا

لا يُقدّر الرد الواحد التوزيع التمثيلي للسكان المستهدفين المحددين. لذلك، لا يمكنه إنتاج درجة GEO إجمالية.

8. وجود الحدث وانتشاره

هذا واحد من أهم التمييزات في هذا القسم.

8.1. وجود الحدث

يجيب على السؤال: "هل حدث هذا التمثيل أو الخطأ مرة واحدة على الأقل؟" قد تكون ملاحظة واحدة صالحة كافية.
التمثيل:

```
I(E) = 1إذا تم ملاحظة الحدث مرة واحدة على الأقل)؛ وإلا 0 = I(E)
```

الرمز 0 هنا لا يعني: "الحدث مستحيل." إنه يشير فقط إلى أنه لم يُلاحظ في القياس الحالي.

8.2. انتشار الحدث

يجيب عن السؤال: "كم مرة يحدث الحدث في السكان المستهدفين المحددين؟" في عينة بسيطة، يمكن حساب المعدل الخام على النحو التالي:

```
p̂ = عدد السجلات الصالحة التي لوحظ فيها الحدث / إجمالي عدد السجلات الصالحة
```

ومع ذلك، في الدراسات الواقعية GEO-1000:

- أوزان السكان،
- طبقات البلد واللغة،
- المستخدمون المطابقون،
- مجموعات اللوحة،
- عدم الاستجابة,
- سجلات غير صالحة،
- تأثير التصميم

يجب أخذها في الاعتبار. لذلك، فإن المعدل الخام ليس دائماً تقدير السكان النهائي.

8.3. لا يمكن تحويل الحدوث والسائدة إلى بعضهما البعض

قد يتم ملاحظة حدث مرة واحدة ولكنه نادر جداً. قد يكون الحدث شائعاً ولكن لم يتم ملاحظته على الإطلاق في تجربة صغيرة. الجملتان التاليتان مختلفتان:

- "تم ملاحظة هذا الخطأ مرة واحدة."
- "تم ملاحظة هذا الخطأ لدى 12% من المستخدمين."

الجملة الثانية تتطلب مقام وعينة محددين.

9. المظهر الإحصائي للتحيز نتيجة الملاحظة الفردية

لقد طرحنا سؤالًا على منتج ذكاء اصطناعي مرة واحدة. وقد استوفى الجواب معيار النجاح. النسبة الخام:

```
p̂ = 1/1= 1= 100%
```

يبدو أنها 100 في المائة. لكن ذلك 100 في المائة لا يعني أن كل عضو من المجموعة المستهدفة سيرى النتيجة نفسها. للتوضيح فقط، يُظهر الجدول أدناه تقريبًا 95 في المائة من فترات الثقة لويلسون تحت الملاحظات المستقلة لبيرنولي والعينة العشوائية البسيطة غير المرجحة: [K09]

تقريب فاصل ويلسون 95%	النسبة الخام	ملاحظة صالحة	الاستجابة السابقة
20.7%-100%	100%	1	1
59.6%-98.2%	90%	10	9
88.8%-97.8%	95%	100	95
93.5%-96.2%	95%	1,000	950

هذا الجدول اصطناعي وتمثيلي. إنه لا يثبت:

- أن هذه الفواصل صالحة لجميع تصاميم أخذ العينات لـ GEO-1000

- أن ملاحظات 1,000 كافية لجميع المجموعات الفرعية من اللغات والدول
- أن التجميع والوزن غير فعالين
- أن نسبة الدقة الخام 95 تعني تلقائيًا الملاءمة

ما يوضحه فعليًا هو:

> مع انخفاض عدد الملاحظات، يزداد الشك في نسبة السكان.

النسبة المئوية 100 التي تُرى في ملاحظة واحدة يمكن أن تكون متوافقة مع نطاق سكاني محتمل واسع جدًا.

10. لماذا "1,000" ليست رقماً سحرياً؟

قاعدة المرشح الأساسية — القاعدة التأسيسية للمرشح

تحمل GEO-1000 هذا الاقتراح التأسيسي:

> لكل منتج ذكاء اصطناعي مراقب ولكل موجة قياس رئيسية، يتم استهداف الحد الأدنى من *1,000* ملاحظات مستخدم صالحة بشكل افتراضي.

هذا الحكم ليس بعد الحد الأدنى العلمي النهائي. سيتم إعادة تقييمه بناءً على نتائج التجربة التجريبية والتطبيق الخارجي.

10.1. لماذا هو بداية قوية؟

بموجب افتراض العينة العشوائية البسيطة، عندما يكون المعدل الحقيقي حوالي 50%، فإن هامش الخطأ النظري 95% لملاحظات 1,000 يكون تقريبًا ±3.1 نقطة. يمكن أن يوفر هذا دقة بداية ذات مغزى على مستوى المنتج العام.

10.2. لماذا لا يحل كل مشكلة؟

ملاحظات 1,000:

- إلى عشرات الدول،
- إلى العديد من اللغات،
- إلى الخطط المجانية والمدفوعة،
- إلى الواجهات على الويب والجوال،
- إلى اللوحات المحكمة والطبيعية

عند التقسيم، تنكمش الخلايا الفرعية. على سبيل المثال، إذا كان هناك فقط 50 ملاحظات لكل من لغات 20، يمكن أن يكون عدم اليقين القائم على اللغة أعلى بكثير من درجة النموذج الإجمالية.

10.3. تأثير التصميم

الملاحظات:

- من نفس لوحة البحث،
- من نفس البلد،
- من نفس المستخدم،
- من تصميم مطابق

موجودة، قد لا ينطبق افتراض العينة المستقلة البسيطة. يمكن أن يكون عدم اليقين الفعلي أوسع.

10.4. منهجية NOMOS

تعمل ملاحظات 1,000 كالتالي:

- نقطة المرجع التأسيسية للبروتوكول,
- الحد الأدنى الافتراضي للبدء على مستوى النموذج,
- عتبة المرشح ليتم معايرتها مع الطيار

NOMOS لا يدعي أن ملاحظات 1,000 كافية تلقائيًا لكل بلد أو لغة أو حالة استخدام.

11. الملاحظة المستقلة ليست هي نفسها الملاحظة المتكررة

يمكن للمستخدم أن يطرح نفس السؤال عشر مرات. هذه العملية يمكن أن تكون ذات قيمة. ومع ذلك، فهي لا تنتج عشرة مستخدمين مستقلين.

11.1. نفس المستخدم - نفس الجلسة تكرار

يمكن قياس:

- تباين تجديد الإجابة,
- الثبات داخل الجلسة،

احتمالية الإجابات البديلة. لا يقيس انتشار السكان بمفرده.

11.2. نفس المستخدم–تكرار جلسة جديدة

قد يقلل من سياق الدردشة السابقة. ومع ذلك:

- نفس الحساب،
- نفس الخطة،
- نفس الدولة،
- نفس التخصيص،
- نفس المستخدم

قد يستمر في خلق اعتماد إحصائي بين الملاحظات.

11.3. اختبار مستخدم واحد لعدة منتجات ذكاء اصطناعي

يمكن أن يساعد هذا التصميم في المقارنة المطابقة عبر النماذج. ومع ذلك، فإن الملاحظات تعتمد على مستوى المستخدم. يجب الحفاظ على هذا الاعتماد عند التسجيل والتحليل.

11.4. مستخدمون مختلفون يستخدمون نفس الحساب

قد يوجد أشخاص مختلفون. ومع ذلك:

- تاريخ الحساب،
- التخصيص،
- الخطة،
- الذاكرة

يمكن مشاركتها. لا يمكن اعتبار هذه السجلات مستقلة تلقائيًا.

11.5. نصّ استقلالية معيارية

يجب تصميم الملاحظات لتكون مستقلة قدر الإمكان؛ يجب تسجيل واعتماد المستخدم المعروف، الحساب، الجلسة، اللوحة، الدولة، الجهاز، والاعتمادات المتكررة وأخذها في الاعتبار في التحليل.

لا يجب أن يتم اختزال الاستقلالية إلى ثنائي:

- نعم،
- لا

إذا لزم الأمر، استخدم أحد حالات الاعتماد التالية:

مستقل

مزود

مجمّع

متكرر ضمن المستخدم

تابع

مجهول

12. مبدأ أول مخرَج مناسب

في لوحة السكان الرئيسية، مهمة المشارك ليست العثور على الإجابة الأكثر إيجابية. يجب على المشارك تسجيل أول إخراج مناسب مكتمل تم توليده تحت الشرط المحدد مسبقًا. المشارك:

- لا يمكنه تحديث الإجابة،
- لا يمكنه طلب إجابة جديدة،
- لا يمكنه اختيار الإجابة الإيجابية،
- لا يمكنه اقتصار التعديل على الجزء الجيد فقط،

لا يمكنه إعادة كتابة الإجابة. يمكن أيضًا دراسة سلوك الاستنساخ بشكل منفصل. ومع ذلك: "لقد طرحنا نفس السؤال خمس مرات واخترنا أفضل إجابة." ليست هذه هي نتيجة السكان الرئيسيين. هذا هو اختبار الإخراج الأفضل المحدد. يجب الإبلاغ عنه باسم وطريقة منفصلين.

13. استثناء حدث حرج واحد

لا يمكن لإجابة واحدة تحديد درجة المجتمع. ومع ذلك، هذا لا يعني أن الإجابة الواحدة لا يمكن أن يكون لها تأثير كبير على اتخاذ القرار. NOMOS يحدد مسارين تقييميين منفصلين.

13.1. مسار تمثيل المجتمع

يسأل السؤال: "كم مرة يحدث هذا الناتج في السكان المستهدفين؟" يتطلب عينة كبيرة وصحيحة.

13.2. مسار الحدث الحرج

يسأل السؤال: "ما مدى الضرر المتوقع عند حدوث هذا الناتج مرة واحدة؟" يمكن أن تؤدي نتيجة واحدة مؤكدة إلى تقييم. على سبيل المثال:

- منح رخصة صحية غير موجودة
- إجراء إحالة صحية طارئة خطيرة
- الإبلاغ عن سلطة قانونية غير صحيحة
- إزالة الحدود الأمنية
- خَلط شركة مع كيان قانوني آخر
- تكرار تعليمات تManipulative مخفية عن الناس كما لو كانت نصيحة

يمكن أن يكون حدثًا حرجًا مفردًا.

13.3. فصل القرار بين النتيجة والامتثال

يمكن أن يكون الحدث الحرج عادةً ملاحظة واحدة. على سبيل المثال:

```
1/1000=0.1%
```

قد يبدو هذا المعدل صغيرًا. ومع ذلك، إذا كان الضرر المحتمل للحدث شديدًا، فقد يخلق فجوة لا يمكن إصلاحها في قرار الامتثال. لذلك:

```
يقيس معدل السكان تكرار الحدث؛ بينما يقيس مراجعة الحدث الحرج تأثير الحدث على الضرر والنزاهة.
```

لا يمكن للدقة العالية في المتوسط أن تلغي تلقائيًا الخطأ الفردي الكبير. سيتم تحديد الظروف النهائية للتصنيف الحرج في الفصول اللاحقة من الكتاب.

14. سلم الادعاءات الملاحظة

في هذا القسم، يتم فصل المستويات الأربعة الأولى من الملاحظة فقط بشكل معياري. على المستويات الأعلى:

- الدرجة الموزونة حسب السكان،
- درجة النماذج المتعددة،
- استقرار الزمن،
- تأثير التدخل،
- القيمة المستدامة

سيتم تعريفه في الأقسام التالية.

O-0— لم يتم اختباره

لم يتم قياس منتج الذكاء الاصطناعي أو الحالة ذات الصلة. الحالة الصحيحة:

لم يتم اختباره

جملة لا يمكن الدفاع عنها: "هذا مناسب لأن المشكلة لم تُلاحَظ."

O-1— ملاحظة حدث فردي

هناك نتيجة واحدة صالحة. صياغة يمكن الدفاع عنها: "تحت الظروف المحددة، لوحظ هذا الناتج مرة واحدة." جملة لا يمكن الدفاع عنها: "النظام يمثل الوجود بهذه الطريقة."

O-2— تكرار من المستخدم أو الجلسة

توجد نتائج متعددة لنفس المستخدم أو الحساب أو عائلة الجلسة. صياغة يمكن الدفاع عنها: "في ظروف المستخدم والجلسة المحددة، حدث هذا الناتج في سبع مرات من أصل عشر تكرارات." جملة لا يمكن الدفاع عنها: "سبعة من كل عشرة مستخدمين مستقلين شاهدوا هذا الناتج."

O-3— قياس خلية واحدة متعددة المستخدمين

نفس الشيء:

- منتج الذكاء الاصطناعي،
- واجهة المستخدم
- البلد،
- اللغة،
- حالة اللوحة،
- نسخة المُدخل،
- موجة القياس

يحتوي على العديد من ملاحظات المستخدم المناسبة. صياغة يمكن الدفاع عنها: "78% من 200 الملاحظات الصالحة في الخلية المسيطر عليها في تركيا/Turkish استوفت معيار الانتقال المحدد." لا يمكن تعميم هذه النتيجة تلقائيًا على دول أو لغات أو واجهات مستخدم أخرى.

O-4— القياس متعدد الطبقات عبر الخلايا

توجد عينات محددة مسبقًا من دول ولغات أو خلايا مستخدم مختلفة. على هذا المستوى:

- نتائج الخلية،
- النتيجة المرجحة،
- النطاق،
- عدم اليقين

يمكن الإبلاغ عنها. القواعد التفصيلية لأوزان السكان وعدالة المجموعات الفرعية مذكورة في الفصول اللاحقة من الكتاب.

15. عرض توضيحي اصطناعي APPLE.COM

عرض توضيحي للمنهجية الاصطناعية / المستخدمون، منتجات الذكاء الاصطناعي، الردود، المعدلات، والنتائج أدناه كلها خيالية. ولا تمثل الأداء الفعلي لشركة .Apple Inc أو سلوك أي منتج ذكاء اصطناعي حقيقي.

15.1. السيناريو A — استجابة إيجابية واحدة

يقدّم مستخدم اصطناعي السؤال الآتي إلى منتج الذكاء الاصطناعي المُحاكى A: «إلى أي نوع من الشركات تنتمي Apple.com؟» الاستجابة الاصطناعية: «Apple شركة تكنولوجيا مقرها الولايات المتحدة، تطوّر إلكترونيات استهلاكية وبرمجيات وخدمات رقمية». التقييم الأولي:

- هوية الكيان صحيحة
- الفئة الرئيسية صحيحة
- لا توجد أخطاء حرجة واضحة
- التمثيل الأساسي مقبول

النتيجة الخام:

```
1/1=100%
```

تصريح عام غير صحيح: "المنتج الذكاء الاصطناعي أ يمثل شركة أبل بدقة .100%" التصريح العام الصحيح: "في العينة التركيبية، الملاحظة الفردية المحددة استوفت معيار الانتقال. لم يتم تحديد انتشار السكان وموثوقية المنتج."

15.2. السيناريو ب — الاكتشاف بعشر ملاحظات

عشرة مستخدمين تركيبيين مستقلين يستخدمون نفس إصدار الطلب الدلالي. النتيجة التركيبية:

- 7 تجاوبيات تم اجتيازها
- استجابة 1 صحيحة جزئياً
- استجابة 1 تحتوي على معلومات قديمة
- استجابة 1 تعيّن النشاط الرئيسي الخاطئ

معدل النجاح الخام:

```
7/10=70%
```

تحت افتراض بسيط ومثالي للعينات المستقلة، فإن فترة ويلسون التقريبية لـ 95 بالمئة بالضرورة واسعة. النتيجة تبقى استكشافية لأنها تعتمد على:

- عينة صغيرة،
- خلية واحدة،
- موجة واحدة

الحالة الصحيحة:

استكشافية — غير كافية للنقاط العالمية المتاحة للعامة لـ *NOMOS*.

15.3. السيناريو C — 1,000 موجة اصطناعية مرصودة

افترض بشكل اصطناعي أن الشروط التالية مستوفاة:

- المنتج الذكائي أ
- 1,000 ملاحظات صالحة
- السكان المستهدفون المحددون مسبقًا
- نسخة صياغة المُدخل المؤمّنة
- طبقات الدولة واللغة المحددة
- الاستجابة الأولية المناسبة غير المعدلة
- سجلات الأدلة الكاملة
- قواعد التحكيم المحددة مسبقًا

النتيجة الاصطناعية:

حالة الاستجابة	الرقم
استيفاء معيار الانتقال	918
صحيح جزئيًا	37
ادعاء بمادة غير مدعومة	23
معلومات قديمة	12
مرشح الخطأ الحرج	6
غير قابل للتصنيف	4
الإجمالي	1,000

معدل النجاح الخام البسيط:

```
918/1000=91.8%
```

الفاصل الزمني التقريبي 95% ويلسون تحت الافتراض المثالي غير الموزون: هو. ومع ذلك، هذا الرقم وحده ليس بعد النتيجة النهائية لـ GEO NOMOS. لأنه هناك عوامل إضافية يجب تقييمها:

- أوزان البلد واللغة
- تأثير التصميم
- أدنى نتيجة للغة
- معدل الملاحظات غير الصالحة
- اتفاق المقيمين
- التحقق من مرشحي الأخطاء الحرجة
- مجموعات المستخدمين غير المختبرة
- الوقت و واجهة المستخدم كحدود للقياس

يجب أن يخضع ستة مرشحين للأخطاء الحرجة أيضًا لمراجعة حادث منفصلة. التقرير الصحيح: 'معدل تمرير الاستجابة الخام في الموجة الاصطناعية هو 91.8 بالمئة. تصف هذه النتيجة الملاحظات قبل تطبيق أوزان السكان وتأثيرات التصميم. نظرًا لأنه تم تحديد ستة مرشحين للأخطاء الحرجة، فإن قرار الامتثال يتطلب مراجعة منفصلة.' التقرير غير الصحيح: 'درجة GEO لشركة NOMOS Apple هي 91.8 وهي متوافقة.' لا يمكن نسب البيانات الاصطناعية بهذه الطريقة إلى شركة حقيقية أو نظام حقيقي.

16. الأحكام الإلزامية المعيارية

N-01

لا يمكن استخدام مخرج نظام توليدي واحد كدرجة عامة للجيو.

N-02

نتاج مفرد يُلاحظ بمفرده:

- منتج الذكاء الاصطناعي و واجهة المستخدم
- البلد،
- اللغة،
- مُدخل,
- الجلسة،
- الوقت

يجب الإبلاغ عنه مع ظروفه.

N-03

مخرج فردي:

- نموذجي
- مستمر
- عالمي
- ينتمي إلى جميع المستخدمين

لا يمكن أن يُعرض كسلوك.

N-04

لا يمكن احتساب التكرارات من نفس المستخدم أو الحساب أو الجلسة كملاحظات سكانية مستقلة دون الكشف عن الاعتماد.

N-05

في لوحة السكان الرئيسية، يجب على المشارك تسجيل أول مخرج مناسب محدد مسبقًا؛ لا يمكن إجراء الاستنساخ لاختيار استجابة إيجابية.

N-06

يجب تسجيل رفض النظام، رسالة الخطأ، أو عدم القدرة على الاستجابة كنتيجة نظام صالحة إذا تم تطبيق البروتوكول بشكل صحيح.

N-07

لا يمكن تضمين الإجابات المقلمة أو المعدلة أو المعاد كتابتها يدويًا أو الإجابات التي لا يمكن التحقق من سياق الطلب الخاص بها في حساب النتيجة الرئيسية.

N-08

يُحظر اختيار لقطات الشاشة الإيجابية فقط وامتناع عن توزيع النتائج.

N-09

لا يمكن استخدام الملاحظات الاصطناعية في درجات المستخدم الحقيقي أو النظام الحقيقي.

N-10

لا ينبغي نشر نسبة السكان بدون التغطية والمقام وشرح مناسب لدرجة عدم اليقين.

N-11

لا يمكن تعويض احتمال حدوث خطأ حرج بهدوء ضمن المتوسط الكلي.

N-12

يمكن تطبيق نتيجة الملاحظة فقط على الكون الاستنتاجي الذي تم تعريفه بشكل فريد؛ لا يمكن تعميمها على الدول أو اللغات أو المنتجات أو الفترات الزمنية غير المقاسة.

N-13

يجب أن يكون للقياس أو التصنيف أو البيان العام صاحب مسؤول بشري أو مؤسسي.

N-14

السجلات المجهولة أو غير المختبرة أو غير الصالحة لا يمكن تحويلها بهدوء إلى نتيجة إيجابية.

17. المرشح GEO-1000 القاعدة الرئيسية

المرشح-001

بالنسبة لكل منتج ذكاء اصطناعي يتم تدقيقه ولكل موجة قياس للسكان الرئيسيين، يجب استهداف الحد الأدنى من *1,000* ملاحظات مستخدم صالحة بشكل افتراضي.

هذه الجملة المرشحة:

- دقة توقع المنتج العام،
- التكلفة،
- قابلية التطبيق،
- العدالة بين الدول واللغات

هذا يخلق توازنًا أوليًا بين هذه الاعتبارات. ومع ذلك، لا يضمن تغطية كافية لـ:

- جميع درجات الدول،
- جميع درجات اللغات،
- جميع فئات المستخدمين الفرعية

يجب بدلاً من ذلك اختبار الكفاية من خلال:

- التجارب التجريبية،
- التكرارات المستقلة،
- تأثير التصميم،
- حساسية الفئات الفرعية

تحدد هذه التقييمات الحالة المعيارية النهائية.

18. أشكال الفشل

F-01— لقطة الشاشة المختارة

يتم نشر الاستجابة الأكثر إيجابية فقط من بين العديد من النتائج.

F-02— التعامل مع مستخدم واحد كالسكان

يتم تعميم تجربة شخص واحد على كامل الكون المستهدف.

F-03— عد تكرارات نفس المستخدم كأشخاص جدد

التكرارات التي يولدها نفس الشخص أو الحساب يتم الإبلاغ عنها كمشاركين مستقلين.

F-04— تجديد الاستجابة واختيار الأنسب

يتم تجاهل الاستجابات الأولية أو السلبية؛ ويتم حفظ الاستجابة الأنسب كنتيجة للقياس.

F-05— انحراف صياغة المُدخل

يقوم المشاركون بتغيير الكلمة أو النبرة أو نطاق صياغة المُدخل؛ وتُدمج النتائج كما لو كانت من نفس التجربة.

F-06— تلوث السياق

المحادثات السابقة تؤثر على الاستجابة؛ ويتم عرض السجل كجلسة نظيفة.

F-07— إخفاء التخصيص

الذاكرة، التعليمات الخاصة، أو تاريخ المستخدم مرئي؛ يتم الإبلاغ عن النتيجة كرد عام للمستخدم.

F-08— نموذج ومنتج مربك

يُكتب فقط اسم النموذج العام أو المزود؛ لم يتم تسجيل المنتج الفعلي أو الواجهة أو حالة الأداة.

—F-09 إزالة الأخطاء والعيوب من المقام

تم الحذف لجعل النتائج تبدو إيجابية عندما لا يستجيب النظام للسجلات.

—F-10 احتساب الالتقاط غير الصالح كفشل النظام

تم الإبلاغ عنه كخطأ من المشترك، أو قصور الأدلة، أو نتيجة غير صحيحة لمنتج الذكاء الاصطناعي.

—F-11 تعميم لغة وبلد واحد على مستوى العالم

يتم تطبيق نتيجة موقع واحد على مستوى العالم.

—F-12 استخدام البيانات الاصطناعية كأنها حقيقية

يتم إضافة السجلات التي تم إنشاؤها لأغراض العرض إلى الأداء الفعلي.

—F-13 فقدان حدث حاسم في المتوسط

يتم جعل الأخطاء الشديدة في الصحة أو القوانين أو التراخيص أو الأمان غير مرئية ضمن متوسط مرتفع.

—F-14 استنتاج تأثير التدخل من نجاح واحد

يتم اعتبار الاستجابة الإيجابية الوحيدة التي تم تلقيها بعد تدخل GEO على أنها التأثير السببي للدراسة.

—F-15 الحكم على النظام بأكمله من خطأ واحد

يتم تعميم خطأ واحد كما لو أن منتج الذكاء الاصطناعي خاطئ دائمًا وخاطئ لجميع المستخدمين.

—F-16 تحديد السكان المستهدفين بعد ذلك

بعد رؤية النتيجة، يتم إعلان مجموعة المستخدمين ذات النجاح العالي على أنها "السكان المستهدفون الفعليون".

19. إجراء التدقيق

يجب على المدقق اتباع التسلسل أدناه عند تقييم صحة ملاحظة واحدة أو ادعاء GEO مبكر.

الخطوة —1 تسجيل الادعاء العام

يتم تحديد الجملة المقترحة بالكامل. مثال: "ChatGPT يعرف شركتنا بشكل صحيح."

الخطوة —2 تحديد مستوى الادعاء

قد تشير الجملة إلى:

- حدث واحد،
- الانتشار،
- الاستقرار،
- العواقب العالمية،
- السببية،
- التأثير التجاري

يجب تحديد المستوى المناسب.

الخطوة —3 تحديد عدد الملاحظات الداعمة

كم عدد المستخدمين الحقيقيين؟ كم عدد الحسابات؟ كم عدد الجلسات المستقلة؟ كم عدد التكرارات؟ كم عدد منتجات الذكاء الاصطناعي؟ كم عدد اللغات والدول؟

الخطوة 4— التحقق من شروط الملاحظة

هل تم تسجيل منتج الذكاء الاصطناعي، واجهة المستخدم، المُدخل، الجلسة، الدولة، اللغة والوقت جميعها؟

الخطوة 5— فحص سلامة الأدلة

هل هناك استجابة كاملة؟ هل هناك لقطة شاشة كاملة؟ هل تم اختيار الاستجابة أو تحديثها؟ هل تم تعديل الملف؟ هل السجل اصطناعي؟

الخطوة 6— تحديد هيكل الاعتماد

هل هناك تكرارات لنفس المستخدم؟ هل تم استخدام حساب مشترك؟ هل هناك تصميم نموذج معكوس؟ هل هناك تجميع من نفس اللوحة؟

الخطوة 7— فصل خطأ جمع البيانات عن النتيجة النظامية الصالحة

هل الرفض أو خطأ النظام هو النتيجة الفعلية؟ أم أن المشارك انتهك البروتوكول؟

الخطوة 8— قارن الادعاء بمستوى الدليل

هل تم استخدام ملاحظة واحدة لاستنتاجات على مستوى السكان أو العالمية؟

الخطوة 9— افتح المسار الحرج

إذا كان رد واحد يحمل ضررًا شديدًا متوقعًا، يتم بدء تحقيق حادثة منفصل.

الخطوة 10— حدد القرار

يتم تقليل الحكم العام فقط إلى المستوى الذي يحمله الدليل.

20. الدليل الضروري

بالنسبة لملاحظة أو ادعاء ضمن نطاق هذا القسم، يجب السعي للحصول على السجلات التالية بالقدر الذي تكون فيه ذات صلة:

- هوية الكيان المدقق
- الإفصاح الكامل عن ادعاء GEO للجمهور
- عدد الملاحظات
- عدد المشاركين الفريدين
- منتج الذكاء الاصطناعي و واجهة المستخدم
- النموذج أو الإصدار المرئي، إذا كان متاحًا
- نوع الخطة أو الحساب
- الدولة والموقع
- معرّف النص، الإصدار، والنص الكامل
- حالة الجلسة والتخصيص
- طابع زمني UTC
- موجة القياس
- الاستجابة الخام الكاملة
- لقطة شاشة أو دليل مكافئ

- حالة إعادة الإنتاج أو اختيار الإجابة
- حالة صلاحية الملاحظة
- سجل الاعتماد والمطابقة
- حالة تركيبية أم حقيقية
- سجل الملاحظة غير الصالحة
- سجل الحدث الحرج
- نطاق النتيجة
- بيان عدم اليقين
- الشخص أو المؤسسة المسؤولة

قد لا تكون كل هذه السجلات مطلوبة بنفس التفصيل لكل ملاحظة اكتشاف منخفض المخاطر. ومع ذلك، إذا كان من المقرر منح درجة GEO واسعة أو بند مطابقة للجمهور، فلا يمكن ترك الحقول الأساسية فارغة.

21. قائمة تدقيق التدقيق

هل تم تحديد الكيان المدقق بالكامل؟ هل تم تسجيل منتج الذكاء الاصطناعي و واجهة المستخدم؟ هل تم تثبيت النص قبل القياس؟ هل قام المشارك بتغيير النص؟ هل تم معرفة شروط الجلسة والتخصيص؟ هل تم تسجيل معلومات بلد المستخدم والموقع؟ هل تم الحفاظ على الاستجابة الكاملة؟ هل هو الإخراج الأول المناسب أم أفضل إخراج مختار؟ هل يوجد سجل كامل للأدلة؟ هل تم فصل العدد الفعلي للمستخدمين عن عدد الملاحظات؟ هل تم احتساب الملاحظات المكررة والمطابقة بشكل مستقل؟ هل تم التعامل بشكل صحيح مع حالات الرفض وأخطاء النظام في المقام؟ هل تم الاحتفاظ بالسجلات غير الصالحة وتوضيح سببها؟ هل تم فصل البيانات التركيبية والحقيقية؟ هل يتم تقديم استجابة واحدة كنتيجة للسكان؟

هل يُطبَّق الادعاء فقط على المنتج المقاس، أو البلد، أو اللغة، أو الوقت؟ هل تم تقييم احتمال حدوث خطأ جسيم بشكل منفصل؟ هل يحمل الناتج بيانًا بالثقة أو عدم اليقين؟ هل تُعرض المناطق غير المختبرة بوضوح؟ هل يتجاوز البيان العام مستوى الملاحظة الفعلية؟ إذا لم يُجب على أحد الأسئلة الواقعية، فإن الناتج يحمل ثقة محدودة. إذا تُرِك أكثر من سؤال واحد بدون إجابة، فلا ينبغي استخدامه لتقييم GEO العام.

22. الاعتراضات والردود

الاعتراض —1 "المستخدم الفعلي يرى بالفعل إجابة واحدة فقط."

هذا صحيح. الإجابة الوحيدة التي يراها المستخدم الفردي هي تجربة ذلك الشخص الحقيقية. لذلك، يتم تسجيل الأحداث الفردية. ومع ذلك، فإن الخبرة الفردية ليست هي نفسها خبرة السكان. نظام NOMOS يحافظ على كلا الحقائق معًا:

> تجربة مستخدم واحد حقيقية. / المستخدم الواحد ليس هو جميع السكان.

اعتراض —2 "إذا كان النظام يعطي نفس الإجابة في كل مرة، أليست إجابة واحدة كافية؟"

الادعاء بأن النظام يعطي نفس الإجابة في كل مرة يحتاج أيضًا إلى قياس. بالإضافة إلى ذلك:

- الوقت،
- الاسترجاع،
- البلد،
- اللغة،

- الخطة،
- التخصيص،
- واجهة المستخدم

قد يختلف. حتى النظام الذي يُفترض أنه حتمي لا يمكن تعميمه خارج الشروط المحددة.

اعتراض — 3 "لماذا بالضبط 1,000؟"

الرقم 1,000 ليس حدًا ساحرًا لليقين. GEO-1000 يعمل كنقطة مرجعية تأسيسية توازن بين:

- حجم مستخدمين كبير،
- تقدير واسع على مستوى المنتج،
- تكلفة قابلة للتطبيق،
- سهولة الفهم العامة.

قد تتطلب المجموعات الفرعية المزيد من الملاحظات. يجب أن تتحقق الدراسات التجريبية من الحد أو تعدله.

اعتراض — 4 "دع نفس 1,000 الأشخاص يختبرون جميع منتجات الذكاء الاصطناعي."

يمكن أن تكون هذه الطريقة مفيدة للمقارنة الزوجية بين النماذج. ومع ذلك، قد يؤدي اختيار المستخدمين المتقدمين فقط لذكاء اصطناعي والذين يمكنهم الوصول إلى جميع المنتجات إلى تحيز في السكان المستهدفين. أيضًا، الملاحظات من نفس الشخص تعتمد على بعضها البعض. سيتم مقارنة التصاميم المزدوجة بالكامل، والمزدوجة جزئيًا، والمتوازنة لاحقًا.

اعتراض — 5 "أليس لقطة الشاشة دليلاً كافياً بالفعل؟"

لقطات الشاشة مهمة. ومع ذلك، بمفردها لا تظهر دائماً ما يلي:

- الزمن الحقيقي
- دولة
- خطة
- المحادثة السابقة
- التخصيص
- تحديث الإجابة
- تغيير الملف
- الرد الكامل

لهذا السبب، تُعَد لقطة الشاشة جزءًا مهمًا من حزمة الأدلة. إنها ليست كل شيء.

الاعتراض — 6 "لماذا ندرج نظامًا لا يستجيب في النتيجة؟"

قد تكون تجربة المستخدم الحقيقية هي عدم الحصول على إجابة. الرفض أو الخطأ:

- فني،
- أمني،
- سياسي،
- الوصول

قد يكون هو النتيجة. إذا تم تطبيق البروتوكول بشكل صحيح، يجب تسجيل الحدث. يُصنَّف عدم استجابة النظام بسبب خطأ فني من قبل المشارك بشكل منفصل.

اعتراض 7— "لماذا يجب أن تؤثر خطأ شديد واحد على التفتيش بأكمله؟"

خطأ واحد لا يثبت أن معظم السكان ارتكبوا نفس الخطأ. ومع ذلك، فإن الضرر المحتمل لخطأ واحد يمكن أن يكون شديدًا. لذلك:

- تكرار النتيجة
- القرار الحرج والضرر والنزاهة

يتم تقييمها بشكل منفصل. المتوسط لا يمكنه مسح الضرر البشري تلقائيًا.

اعتراض 8— «إذا لم نستطع قول أي شيء برد واحد، ما هي قيمة الاختبارات المبكرة؟»

الاختبارات الفردية وعينة صغيرة:

- توليد الفرضيات
- اكتشف أنواع الأخطاء
- اختبارات التجارب،
- اختبر نظام الالتقاط،
- قم معايرة القضاة،

يمكن أن تكشف عن أحداث حرجة. ما هو الخطأ هو عدم إجراء اختبار استكشافي. ما هو الخطأ هو تقديم نتيجة الاختبار الاستكشافي كدرجة سكانية.

المبدأ الشائع للفصل 24

تم كتابة هذا الفصل للقضاء على مفهوم خاطئ واحد: "لقد سألت الذكاء الاصطناعي، وحصلت على إجابة؛ هذه هي نتيجة GEO الخاصة بي." لا. أنت لم تحصل بعد على نتيجة GEO عامة. ما لديك هو ملاحظة. هذه الملاحظة:

- حقيقية،
- هامة،
- وحتى حرجة

قد يكون كذلك. ومع ذلك، لا يزال مجرد ملاحظة واحدة. للحصول على درجة GEO، وجود:

- المستخدمين،
- منتجات الذكاء الاصطناعي،
- الدول،
- اللغات،
- الجلسات،
- الأوقات

ما يجب قياسه هو توزيع التمثيل عبر منتجات الذكاء الاصطناعي، والدول، واللغات، والجلسات، والوقت. لقطة شاشة واحدة ليست توزيعاً. عشر تكرارات آلية ليست عشرة أشخاص مستقلين. ولا تمثل ألف إجابة من عينة فقيرة سكان العالم. حتى العينة النسبية للسكان قد تكون غير مكتملة بدون عدالة لغوية. المتوسط العالي لا يمكن أن يمحو الأخطاء الحرجة. ولذلك فإن القانون الأول لقياس NOMOS هو:

> الإجابة الواحدة هي حدث. / الإجابات المتعددة هي نمط. / العينة المحددة هي تقدير للسكان. / التكرار على مدى الزمن هو مقياس للاستقرار. / لكن أيًا من هذه وحدها ليست سببًا أو قيمة مستدامة.

طلب قسم 1 من NOMOS

لا تحضر لي لقطة شاشة واحدة ثم تتحدث عن النظام بأكمله.

لا تجعل إجابة مستخدم واحد هي إجابة سكان العالم.

لا تحوِّل النسخ المكررة من نفس الحساب إلى أشخاص مستقلين.

لا تقم بتحديث الإجابة وعرض المفضل كما لو كان التوزيع الحقيقي.

لا تزيل الرفض أو الأخطاء أو المجهولات من المقام.

لا تعمم لغة واحدة على العالم، أو دولة واحدة على كل الدول، أو منتج ذكاء اصطناعي واحد على جميع الأنظمة التوليدية.

لا تحول نجاحًا واحدًا إلى دليل على تأثير التدخل، أو خطأ واحد إلى صفة ثابتة للنظام بأكمله.

لا تُغرق خطأً شديدًا واحدًا ضمن المتوسط.

أولاً، سجّل الحدث. / ثم قس النمط. / ثم حدد السكان المستهدفين. / ثم أظهر الاعتمادية وعدم اليقين. / وتحدث فقط بالقدر الذي تحمله الأدلة.

جملة ختام الفصل

تمثيل الكيان في أنظمة الذكاء الاصطناعي ليس إجابة واحدة. منتجات الذكاء الاصطناعي هي توزيع يجب قياسه عبر المستخدمين والدول واللغات والجلسات والأوقات.

النواة المعيارية

يجوز لمخرج واحد من نظام توليدي أن يثبت وقوع استجابة بعينها في ظروف موثقة. ولا يجوز استخدام مخرج واحد للاستدلال على: - الانتشار في المجتمع المستهدف، - السلوك المعتاد للمنتج، - الاستقرار عبر الزمن، - الاتساق بين اللغات، - الاتساق بين البلدان، - الاتساق بين المنتجات، - أثر التدخل، - يقين التوصية، - الأثر التجاري، - المطابقة، - أو درجة **GEO**. ويجب أن تستند أي ادعاءات أوسع نطاقًا إلى مجتمع مستهدف محدد مسبقًا، وتصميم صالح متعدد الملاحظات، وبنية تبعية موثقة، وحزمة أدلة، وقاعدة تقييم، وبيان للنطاق، وإفصاح عن عدم اليقين. ويجوز لمخرج واحد ذي ضرر مرتفع أن يفعّل مراجعة مستقلة لحادثة حرجة، من دون أن يُعامل بوصفه دليلًا على مدى الانتشار في المجتمع المستهدف.

الفصل 02

التمثيل هو توزيع

الأسس المنهجية / K04; K05; K09

حد الفصل

أقرت المادة 1 الحكم التالي:

> الاستجابة الواحدة ليست درجة *GEO*. الاستجابة الواحدة هي ملاحظة فردية حدثت في ظروف مسجلة.

يجيب هذا القسم على السؤال التالي:

> إذا لم تكن الاستجابة الواحدة هي درجة *GEO*، فما هو الشيء الفعلي الذي نريد قياسه؟

الإجابة هي:

> تمثيل كيان في الأنظمة التوليدية هو توزيع متعدد الأبعاد يتشكل عبر منتجات الذكاء الاصطناعي والمستخدمين والدول واللغات وظروف الجلسة وصياغات المُدخل والأوقات.

هذا الفصل:

- توزيع التمثيل،
- الأبعاد الأساسية للتوزيع،
- الفرق بين المتوسط والتوزيع
- عدم المساواة بين المجموعات الفرعية،
- الاستقرار،
- مشكلة التغطية والملاحظات المفقودة،
- خطر الأخطاء النادرة والخطيرة،
- الهيكل المفاهيمي لبطاقة النتائج المتاحة للجمهور

يحدد. هذا الفصل لم يتطرق بعد إلى:

- عدد المشاركين الذين سيتم تعيينهم لكل بلد،
- كيفية إنشاء إطار العينة،
- كيفية التحقق من صحة الترجمات السريعة،
- كيفية تقييم الادعاءات الذرية،
- الصيغة الدقيقة للنتيجة النهائية NOMOS،
- عوائد المطابقة

لا يُقرّ هذا الفصل هذه المسائل بشكل معياري؛ تتناول الفصول اللاحقة هذه المسائل بشكل منفصل. قبل تحديد صيغة التقييم، يعرّف الفصل 2 الواقع الذي يُراد للتقييم أن يمثله.

التحدي NOMOS

اعتبر منتجين للذكاء الاصطناعي. كل منهما يحصل على نفس متوسط النقاط العالمي:

90 / 100

نتائج اللغة للمنتج الأول كما يلي:

- الإنجليزية: 90
- التركية: 91
- الألمانية: 89
- اليابانية: 90
- العربية: 90

نتائج اللغة للمنتج الثاني كما يلي:

- الإنجليزية: 99
- التركية: 98
- الألمانية: 97
- اليابانية: 86
- العربية: 70

المتوسط البسيط لكلا المنتجين هو 90. هل لهما نفس جودة GEO؟ لا. المنتج الأول ينتج تمثيلًا مشابهًا بلغات مختلفة. المنتج الثاني ينتج تمثيلًا ممتازًا في بعض اللغات ولكنه ضعيف جدًا في لغة أخرى. يخفي المتوسط الفردي هذا الاختلاف. الآن ضع في اعتبارك منتجين آخرين. 95% من الردود لكلا المنتجين تلبي معايير الاجتياز. يتكون 5% المتبقي في المنتج الأول من:

- نواقص نطاق صغيرة،
- نقص في التوضيح محدود،
- بعض المجهولين

النتائج. يتكون 5% المتبقي في المنتج الثاني من:

- سلطة صحية غير صحيحة.
- رخصة غير موجودة،
- هوية قانونية غير صحيحة،
- نصيحة غير مناسبة

موجودة. معدل الدقة الخام لكليهما مرة أخرى هو 95 بالمئة. هل لهما نفس خطر التمثيل؟ لا. الآن ضع في اعتبارك الزوج الثالث. متوسط كلاهما عبر ثلاث موجات قياس هو 90. الأول:

- الموجة 1: 90
- الموجة 2: 91
- الموجة 3: 89

الثاني:

- الموجة 1: 98
- الموجة 2: 90
- الموجة 3: 82

هل لديهم نفس الاتساق مع مرور الوقت؟ لا. في هذه الحالة، القضية ليست فقط: 'ما هو المتوسط؟' نحتاج أيضًا إلى معرفة ما يلي: كيف يتم توزيع النتائج؟ من هو ممثل تمثيلًا جيدًا، ومن هو ممثل تمثيلًا ضعيفًا؟ أي اللغات والدول متأخرة؟ ما مدى خطورة الأخطاء؟ كم تتغير النتيجة مع مرور الوقت؟ أي المجالات لم تُقاس أبدًا؟ بأي أوزان تم إنشاء المتوسط؟ كم عدد الحقائق المختلفة التي تكمن وراء نفس المتوسط؟ الاستنتاج الأول لهذا القسم هو:

> المتوسط نفسه ليس هو التمثيل نفسه.

ينص حكمه الثاني على:

> قد يكون درجة *GEO* ملخصًا للتوزيع؛ لا يمكن أن يحل محل التوزيع.

1. غرض الفصل

يحدد هذا القسم كيفية فهم تمثيل الكيانات في الأنظمة التوليدية قبل اختزاله إلى رقم واحد. الأهداف الرئيسية له هي:

- لإظهار أن التمثيل ليس "وجهة نظر الذكاء الاصطناعي" الثابتة والفريدة
- لتحديد الشروط التي يعتمد عليها ناتج التمثيل
- لتوضيح أن الثنائية صحيحة-خاطئة وحدها غير كافية
- لفصل الدقة المتوسطة عن عدم المساواة بين المجموعات الفرعية
- لإظهار خطر الخطأ الكبير المخفي ضمن المتوسطات العالية
- لإظهار أن الاستقرار والدقة ليسا نفس المفهوم
- لمنع المناطق غير المقاسة من أن تُملأ بالافتراضات الإيجابية ضمن التوزيع
- لتحضير هيكل درجة مستقبلية لـ NOMOS تكون فردية ولكن ليست محكومًا عليها بأن تكون فردية

في نهاية هذا القسم، يجب أن يكون القارئ قادرًا على الإجابة عن السؤال التالي:

> لماذا لا يقتصر ناتج *GEO* لكيان ما على معدل الدقة المتوسط فقط؟

2. الأحكام المعيارية المركزية

> تمثيل الكيان في الأنظمة التوليدية هو التوزيع الشرطي للاستجابات ونتائج التمثيل المختلفة التي يمكن أن تحدث تحت شروط محددة.

هذا التوزيع:

- ليس "الفكر" المجرد لنموذج واحد
- ليس هذا الرأي العام لجميع أنظمة الذكاء الاصطناعي،
- ليس فقط معدل تكرار الموقع،
- ليس فقط تكرار ذكر اسم العلامة التجارية،
- ليس فقط الدقة المتوسطة،

ليس فقط معدل التوصية. توزيع التمثيل، ضمن بروتوكول معين، يظهر ما يلي:

> أي مستخدم يرى أي نوع من التمثيل على أي منتج ذكاء اصطناعي و واجهة المستخدم، بأي لغة، في أي بلد، باستخدام أي مُدخل، تحت أي شرط جلسة، وفي أي وقت.

لذلك، فإن نتيجة التوزيع تعتمد دائمًا على الأسئلة التالية: أي كيان؟ أي منتج ذكاء اصطناعي؟ أي واجهة المستخدم؟ أي فئة مستهدفة؟ أي بلد؟ أي لغة وموقع؟ أي مُدخل؟ أي شرط جلسة؟ أي موجة قياس؟ أي حزمة حقائق؟ أي قاعدة تقييم؟ عندما يتغير أي من هذه الحقول، قد تتغير التوزيعات أيضًا.

3. ماذا يعني كلمة "توزيع"؟

في هذا الكتاب، لا تشير كلمة التوزيع إلى وجهة نظر خفية أو ميتافيزيقية يُفترض أنها موجودة داخل نظام الذكاء الاصطناعي. كما أنها لا تدعي: NOMOS: "هذا هو التوزيع الحقيقي للشركة في ذهن النموذج." التوزيع هنا عملي وقابل للملاحظة.

> إنه نمط النتائج المختلفة للتمثيل التي يمكن الحصول عليها تحت بروتوكول محدد داخل المجتمع المستهدف وضمن ظروف القياس.

يجب تمييز بين توزيعين.

3.1. توزيع التمثيل المستهدف

إنه التوزيع الذي يحاول القياس التنبؤ به لكنه لا يمكن ملاحظته بالكامل بشكل مباشر. يتم تمثيله كالتالي:

P_E(Y | A, C, L, S, P, T, U)

هنا:

- E: الكيان الذي تم تدقيقه
- Y: نتيجة التمثيل
- الإجابة: منتج الذكاء الاصطناعي واجهة المستخدم
- C: البلد أو الحالة الجغرافية
- L: اللغة والموقع
- S: حالة الجلسة والتخصيص الشخصي
- P: الإدخال
- T: موجة الزمن أو القياس
- U: كون المستخدم المستهدف

يطرح هذا التعبير السؤال: "في كون المستخدم المستهدف المحدد، تحت الظروف المحددة، ما هي النتائج التمثيلية التي تحدث مع أي احتمالات؟" لا يمكن قياس جميع عناصر التوزيع المستهدف. يتم تقديرها من خلال أخذ العينات.

3.2. التوزيع التجريبي المشاهد

إنه توزيع الملاحظات الصالحة التي تم الحصول عليها في القياس الفعلي أو الاصطناعي. تمثيله:

P̂_E

يمكن استخدامه. التوزيع التجريبي ليس هو التوزيع الهدف نفسه. إنه تقدير يتم إنتاجه من خلال عينه، والوزن، وتصميم القياس. قد يعتمد حجم الفرق على:

- تصميم العينة
- اختيار المشاركين
- تغطية الدولة واللغة
- معدل الملاحظات غير الصالحة
- عدم الاستجابة
- خطأ المقيم
- موجة القياس
- اختيار صياغة المُدخل

- ظروف المنتج والواجهة
- الوزن
- التجميع والاعتماد

لذلك، لا يمكن نشر التوزيع التجريبي كـ "الحقيقة المطلقة والثابتة للذكاء الاصطناعي." التعبير الصحيح يجب أن يكون: "التقدير الذي تم الحصول عليه للسكان المستهدفين المحددين وظروف القياس المحددة."

4. فضاء التمثيل

تمثيل الكيان لا يحدث في بعد واحد فقط. كل ملاحظة توجد في فضاء متعدد الأبعاد للظروف. التمثيل الأساسي:

```
X_i= (E_i, A_i, U_i, C_i, L_i, S_i, P_i, T_i)
```

هنا:

- Ei: الكيان
- الذكاء الاصطناعي: منتج الذكاء الاصطناعي واجهة المستخدم
- Ui: المستخدم أو وحدة العينة
- Ci: الدولة والطبقة الجغرافية
- Li: اللغة والإعداد الإقليمي
- Si: الحالة، الخطة، وحالة التخصيص
- Pi: المُدخل أو عائلة المُدخلات
- Ti: الوقت وموجة القياس

الاستجابة المولدة:

```
R_i
```

يمكن عرضها. الحالة التمثيلية الناتجة عن تقييم الاستجابة هي:

```
Y_i= g(R_i, V_E, J)
```

ويمكن اعتبارها كذلك. هنا:

- VE: تم استخدام حزمة الحقيقة المُوثَّقة للكيان
- J: قواعد التقييم والتحكيم
- g: عملية التقييم التي تحول الاستجابة إلى نتائج تمثيلية

معنى هذا التمثيل هو كما يلي: لا يمكن تقييم الاستجابة بمفردها دون الإشارة إلى أي كيان، أي حزمة حقائق، وأي قاعدة تقييم تم فحصها وفقًا لها.

5. النتيجة الممثلة ليست قيمة واحدة

فقط الاستجابة بالكامل:

- أن يُجبر على الانتماء إلى فئات

صحيح،

- هي خاطئة

تقديم تصنيف غالبًا ما يكون غير كافٍ. يمكن أن تحدث الحالات التالية معًا في نفس الاستجابة:

- الهوية الصحيحة

- الفئة الرئيسية الصحيحة
- تغطية الخدمة غير مكتملة
- عدد الموظفين قديم
- ادعاء القيادة غير مدعوم
- الاقتباس الصحيح
- توصية غير مناسبة
- نقص غير حرج في الحدود

لذلك، يمكن اعتبار نتيجة التمثيل كموجه نتيجة. التمثيل المحتمل:

```
Y_i= (y_i^id, y_i^cat, y_i^fact, y_i^scope, y_i^time, y_i^evidence, y_i^recommendation, y_i^harm)
```

هنا:

- yiid: دقة الهوية
- yicat: دقة الفئة والدور
- yifact: دقة الأمور الواقعية
- yiscope: الحفاظ على النطاق والحدود
- yitime: الصلاحية الزمنية
- yievidence: التوافق بين الادعاء والدليل
- yirecommendation: التوصية ومدى ملاءمتها للمستخدم
- yiharm: فئة الضرر الناتج عن الخطأ أو التوجيه الخاطئ

ستُحدد قواعد التسجيل الدقيقة لهذه المجالات في الأقسام التالية. المبدأ الإرشادي هنا هو:

> يجب تقييم الإجابة النموذجية ليس فقط بناءً على ما إذا كانت تحتوي على معلومات صحيحة، بل أيضًا بناءً على لمن، وماذا، وفي أي حدود، وبأي دليل، ولمن يتم توصيلها بشكل مناسب.

6. حالات التمثيل الأساسية

قد لا يكون وضع كل ادعاء ذري والاستجابة النهائية نفسه. يفصل هذا القسم مفهوميًا الحالات الأساسية التالية.

6.1. مدعوم

يتم دعم الادعاء بأدلة كافية ضمن النطاق والوقت المحددين.

6.2. متناقض

يدخل الادعاء في صراع مع السجلات الموثوقة والملائمة.

6.3. غير مدعوم

لا يوجد دعم كافٍ للادعاء. لم يثبت أن الادعاء بالضرورة خطأ. ومع ذلك، لا توجد سجلات كافية لقبوله على أنه صحيح.

6.4. قديم - لم يعد صالحًا

قد يكون الادعاء صحيحًا في الماضي. اعتبارًا من تاريخ القياس، لم يعد صالحًا أو لا يمكن التحقق من دقته الحالية.

6.5. مدعوم جزئيًا — مدعوم جزئيًا

تم دعم نطاق محدد فقط، أو السكان، أو الدولة، أو المنتج، أو وقت الادعاء. قد يكون الرد قد تجاوز حد الدعم.

6.6. خارج النطاق — خارج النطاق

الادعاء غير مرتبط بالكيان الذي تم تدقيقه أو بحزمة الحقيقة. قد تكون قد نُقلت من كيان آخر، منتج، دولة، أو خدمة.

6.7. خلط الكيانات — دمج الكيانات

النظام:

- شركتان مختلفتان،
- العلامة التجارية والكيان القانوني،
- شخص والمؤسسة،
- الكيان القديم والجديد،
- قد خلط بين منظمات مختلفة تحمل نفس الاسم

مع بعضها البعض. هذا الخطأ ليس مجرد خطأ واقعي. إنه فساد في كائن التمثيل.

6.8. حَذْفُ الْمَوَاد — مَعْلُومَات مَفْقُودَة

قد تكون الجمل في الرد صحيحة. ومع ذلك، تم حذف معلومات أساسية من شأنها تغيير قرار المستخدم. مثال: الخدمة متاحة فقط في بعض الدول. الرخصة محدودة في سلطة قضائية معينة. السعر لا يشمل الضريبة. النتيجة مبنية على بيانات الطرف الأول. المنتج غير مناسب لبعض المستخدمين. الحذف يمكن أن يخلق استنتاجًا خاطئًا حتى لو لم تكن هناك جملة خاطئة مباشرة.

6.9. رَفْض — تَرْفِيض

النظام رفض الرد. الرفض:

- مناسب,
- غير ضروري,
- غير متسق,
- مبني على الأمن,
- تقني

قد يحدث. Ret لا يُعتبر تلقائيًا فاشلًا أو ناجحًا. يتم تصنيفه بشكل منفصل.

6.10. لا توجد نتيجة

النظام لم ينتج استجابة قابلة للاستخدام لأسباب تقنية أو تشغيلية. ليس هو نفس Ret.

6.11. مجهول

السجلات المتاحة غير كافية لتصنيف موثوق. لا يمكن تحويل المجهول إلى نتيجة إيجابية.

7. الخصائص الرئيسية للتوزيع

لا يمكن تفسير توزيع تمثيلي بناءً على المتوسط فقط. على الأقل، يجب تقييم الخصائص التالية.

7.1. الاتجاه المركزي

يلخص الدقة العامة أو مستوى التحول في الاستجابات. أمثلة على المقاييس:

- المتوسط
- المتوسط المرجح
- الوسيط
- معدل النجاح

الاتجاه المركزي مهم. ومع ذلك، فهو يظهر فقط مركز التوزيع.

7.2. الانتشار والتباين

يشير إلى مدى انتشار النتائج حول المركز. من بين نظامين لهما نفس المتوسط، قد يكون أحدهما ثابتًا بين المستخدمين، بينما الآخر قد يكون متغيرًا بدرجة كبيرة. التباين:

- المستخدمين،
- الدول،
- اللغات،
- الجلسات،
- موجات القياس

يمكن أيضًا فحصها بينهم.

7.3. شكل التوزيع

هل النتائج متجمعة حول نقطة مركزية واحدة؟ أم أنها مقسمة إلى مجموعتين أو أكثر؟ على سبيل المثال:

- إذا رأى نصف المستخدمين دقة عالية جدًا،
- ورأى النصف الآخر دقة منخفضة جدًا

قد يكون المتوسط متوسطًا. ومع ذلك، فإن تجربة المستخدم النموذجية ليست "دقة متوسطة". إنها تجربتان منفصلتان ومتطرفتان.

7.4. اختلافات الفئات الفرعية

هل ترى مجموعات لغوية أو دولة أو مستخدم معينة نتائج مختلفة بشكل منتظم؟ قد يخفي المتوسط العالمي هذا الاختلاف.

7.5. مخاطر الذيل

هل هناك أخطاء نادرة التردد لكنها شديدة؟ يمكن أن يكون للأحداث النادرة تأثير صغير على المتوسط. يمكن أن يكون تأثيرها على الضرر البشري كبيرًا.

7.6. التغطية

كم من كون المستخدمين الهدف تم قياسه فعليًا؟ يمكن العثور على درجة عالية مع تغطية منخفضة.

7.7. الكتلة المفقودة والمجهولة

ما هي نسبة الاستجابات التي هي:

غير معروف،

لم يتم اختبارها،

رفض، عدم استجابة فنية أو سجل غير صالح؟

7.8. استقرار الزمن

إلى أي مدى يتم الحفاظ على التوزيع عبر موجات القياس المختلفة؟

8. توقيع توزيع ممثل للمرشح NOMOS

مفهوم المرشح

يوصى بأن يتم تحديد نتيجة GEO أولاً بواسطة توقيع التوزيع التالي بدلاً من رقم واحد:

```
RDS(E) = (μ, κ, σ, Δ, τ, ρ, u)
```

يُسمى هذا التمثيل توقيع توزيع التمثيل، أي:

توقيع توزيع التمثيل

المكونات:

μ — مستوى التمثيل العام

الدقة الإجمالية أو التنبؤ بالانتقال وفقًا للأوزان والنطاق المحدد.

κ — التغطية

المدى الذي يتم فيه قياس المستخدم المستهدف والبلد واللغة ومنتج الذكاء الاصطناعي وفترة الزمن بشكل كافٍ.

σ — التغير

انتشار بين المستخدمين والخلايا التجريبية.

Δ — عدم المساواة بين المجموعات الفرعية

الفرق المادي بين اللغة أو البلد أو المنتج أو المجموعات المهمة الأخرى.

تاو' — الذيل ومخاطر الخطأ الحرج

معدل وطبيعة الأخطاء النادرة والشديدة.

ρ — الاستقرار الزمني

مرونة التوزيع عبر الموجات.

u — كتلة مجهولة وغير محلولة

غير معروف هو الجزء من التوزيع الذي يظل دون حل أو لم يتم تقييمه بشكل كافٍ. هذه الإشارة ليست بعد صيغة تقييم نهائية. العناصر التالية تحتاج إلى تطوير إضافي:

- تعريف محدد،
- تقدير،
- وزن،
- العرض العام

سيتم الانتهاء منها من خلال الدراسات التجريبية والفصول اللاحقة. المبدأ الأساسي هو:

> يجب أن يجيب ناتج *GEO* ليس فقط على سؤال 'ما مدى الدقة؟' بل أيضًا 'لمن، ما مدى التناسق، ما مدى العدالة، ما مدى الشمولية، وما مدى المخاطرة؟'

9. الغسيل المتوسط

إذا كان المؤسَّسة تُظهِر فقط متوسطًا عالميًا مرتفعًا وتُخفي العيوب الجوهرية للتوزيع، يُسمَّى هذا:

غسل المتوسط

يمكن أن يحدث غسل المتوسط بالطرق التالية:

- إخفاء أضعف لغة
- استبعاد الدول الصغيرة من التقرير
- خلط الأخطاء الحرجة في معدل الخطأ العام
- إزالة نتائج غير المعروفة من المقام
- شمل منتجات الذكاء الاصطناعي ذات الأداء الجيد فقط
- استبعاد موجة القياس الضعيفة
- جعل السكان خارج النطاق غير مرئيين
- تغيير الأوزان بعد رؤية النتائج

يمكن حساب المتوسط رياضيًا بشكل صحيح. ومع ذلك، فإن التمثيل الذي يقدمه للجمهور قد يظل مضللًا. لذلك:

الحساب الرياضي الصحيح للمتوسط لا يعني أنه كافٍ لاتخاذ القرار.

10. واقع مختلف لنفس المتوسط

يمكن أن يكون المتوسط 90 في كل التوزيعات التركيبية التالية.

10.1. توزيع متجانس

جميع المجموعات الفرعية الرئيسية تقع في نطاق 88–92. هذا التوزيع متوازن نسبيًا.

10.2. توزيع مستقطب

يرى نصف المستخدمين نتائج قريبة من 100، بينما يرى النصف الآخر نتائج قريبة من 80. المتوسط هو 90. ومع ذلك، تجربة المستخدم ليست متساوية.

10.3. توزيع مجزأ لغويًا

النتائج جيدة جدًا للغات ذات كثافة سكانية عالية أو موارد كبيرة، لكنها ضعيفة بشكل ملحوظ للغات ذات الموارد القليلة. قد يبدو المتوسط العالمي جيدًا. العدالة اللغوية ضعيفة.

10.4. توزيع مجزأ قائم على الدول

حتى داخل نفس اللغة، قد تختلف النتائج بسبب اختلاف الدول، أو المناطق، أو الوصول إلى المنتجات، أو قاعدة المستخدمين.

10.5. توزيع متقلب زمنيًا

تحدث نتائج مرتفعة جدًا في موجة واحدة، ونتائج منخفضة جدًا في موجة أخرى. حتى لو كان المتوسط مرتفعًا، من الصعب تحديد توقعات موثوقة للمستخدم.

10.6. التوزيع مع مخاطر الذيل

معظم الإجابات صحيحة. ومع ذلك، تحتوي نسبة صغيرة منها على أخطاء صحية أو قانونية أو ترخيصية أو تتعلق بالسلامة.

10.7. توزيع كتلة عالية غير معروفة

معظم الإجابات التي يمكن تقييمها صحيحة. ومع ذلك، فإن جزءًا كبيرًا من السكان المستهدفين أو الإجابات هو:

- مُختبر،
- غير مُختبر،

مجهول

أو غير حاسم. قد يؤدي نشر معدل مرتفع بناءً فقط على الإجابات القابلة للتقييم إلى إخفاء النطاق.

11. تفاوت المجموعات الفرعية

قد لا يمثل المجموع العام التجربة في المجموعات الفرعية. قد تشمل المجموعات الفرعية الرئيسية ما يلي:

- منتج الذكاء الاصطناعي و واجهة المستخدم
- البلد
- اللغة
- الموقع
- الخطة
- الويب أو الجوال
- لوحة مُتحكم بها أو طبيعية
- موجة زمنية
- نية المستخدم
- عائلة الجهاز

قد لا تكون كل هذه المجموعات بنفس الأهمية في كل تدقيق. يجب تحديد المجموعات المادية عند تثبيت النطاق.

11.1. اللغة ليست هي نفسها البلد

متحدثون بالتركية:

- المستخدم في تركيا,
- المستخدم في ألمانيا،
- المستخدم في بلد آخر

يمكنهم استخدام نفس اللغة. ومع ذلك:

- الوصول إلى المنتج،
- السعر،
- الموارد،
- الواجهة المحلية،
- السياق القانوني،
- التخصيص

قد تكون مختلفة. لذلك، لا يمكن شرح جميع الاختلافات بالاعتماد على اللغة أو البلد وحدهما.

11.2. الخلايا المشتركة

تظهر بعض الاختلافات المادية فقط عند فحص مناطق متعددة معًا. المثال: ألمانيا × التركية × الهاتف المحمول × الخطة المجانية × عينة المستخدمين في ظروف الاستخدام الطبيعية قد يكون في حد ذاته خلية تجريبية. ومع ذلك، مع تزايد عدد الخلايا، ينخفض حجم العينة. لذلك، لا يمكن نشر نتيجة حاسمة لكل خلية صغيرة. تصميم القياس: يجب أن يحافظ على المجموعات الفرعية ذات المعنى ولا يجب أن ينتج دقة إحصائية زائفة.

11.3. قاعدة المجموعة الفرعية

مفهوم المرشح

بالإضافة إلى المتوسط العالمي، يُوصى بالإبلاغ عن نتيجة أضعف مجموعة فرعية تم أخذ عينات كافية لها أو أدنى مجموعة 10% بشكل منفصل. يمكن الإشارة إلى ذلك باسم:

قاعدة التمثيل

أو:

الحد الأدنى للغة/الدولة

ومع ذلك، إذا جاء "أدنى درجة" من خلية صغيرة وغير مستقرة، فقد ينتج نتائج غير صحيحة. لذلك، يجب تعريف المقياس الأساسي مع:

- الحد الأدنى للعينة،
- فاصل الثقة،
- جودة الخلية

سيتم تحديد القاعدة الدقيقة في الأقسام التالية.

12. إنه جزء من توزيع النطاق

يتكون التوزيع ليس فقط من الاستجابات المرصودة. ما لم يُرصد أيضًا يحدد معنى النتيجة. على سبيل المثال، تم اختبار الإنجليزية والتركية؛ لم يتم اختبار اليابانية والعربية. تم اختبار واجهة الويب؛ لم يتم اختبار واجهة الهاتف المحمول والخطة المدفوعة. نتائج الإنجليزية والتركية القوية لا تبرر البيان: «التمثيل قوي عبر جميع اللغات والواجهات المستخدمين». الاستنتاج المبرر هو: «تم إجراء القياسات على الواجهات المحددة للويب باللغتين الإنجليزية والتركية. لم يتم اختبار اللغات والواجهات الأخرى للهواتف المحمولة.»

12.1. خريطة التغطية

ينبغي أن تحمل كل توزيعة GEO الخريطة التالية بقدر صلتها:

- الكون المستهدف
- إطار العينة
- الدول المغطاة
- الدول غير المغطاة
- اللغات المغطاة
- اللغات غير المغطاة
- منتجات الذكاء الاصطناعي الخاضعة للمراقبة
- المنتجات غير الخاضعة للمراقبة

- الواجهات المستخدمين الخاضعة للمراقبة
- الواجهات غير الخاضعة للمراقبة
- موجات القياس
- عائلات صياغات المُدخل خارج النطاق

12.2. معدل التغطية

مفهوم المرشح

يمكن تطوير مقياس للتغطية لإظهار النسبة المئوية من الكون المستهدف الممثل بقياس كافٍ. ومع ذلك، فإن التغطية:

- فقط عدد الدول
- فقط عدد اللغات،
- فقط نسبة السكان

لا يمكن قياسها. على سبيل المثال، قد يتم تغطية دول .150 ومع ذلك، قد يكون 95% من المشاركين قد جاءوا من خمس دول فقط. أو قد يتم تغطية 90% من السكان، لكن اللغات الصغيرة قد تُستثنى تمامًا. لذلك، يجب عرض تغطية السكان بشكل منفصل عن تغطية المراقب للدولة/اللغة.

13. الملاحظة المفقودة وعدم الاستجابة

البيانات المفقودة ليست مجرد مشكلة تقنية. يمكن أن يتم تشويه توزيعها بشكل منهجي. الأسباب التي قد تجعل الملاحظة مفقودة قد تشمل:

- فشل المشارك في إكمال المهمة
- المنتج غير متاح في البلد
- قيود الحوسبة
- خطأ تقني
- رفض النظام
- لقطة شاشة غير صالحة
- المشارك يغير مطالبه
- المنتج لا يعمل بلغة معينة
- عدم قدرة المحكمين على اتخاذ قرارات موثوقة

كل سبب من هذه الأسباب يتطلب حالة منفصلة.

13.1. قد لا تكون البيانات المفقودة عشوائية

على سبيل المثال، إذا لم يكن من الممكن إكمال المزيد من المهام في البلدان ذات جودة الاتصال المنخفضة، فإن استخدام التسجيلات الناجحة فقط يمكن أن يحوّل النتيجة العالمية لصالح المستخدمين الأثرياء ذوي الاتصال القوي. إذا ارتفع معدل UNKNOWN بسبب صعوبة تقييم الإجابات بلغة معينة، فقد تخفي الدرجة المحسوبة من الإجابات المعروفة فقط عدم المساواة بين اللغات. لذلك:

> يجب أن يكون معدل البيانات المفقودة وتقرير توزيع سبب فقدان البيانات جزءًا من التقرير.

13.2. المقام — سلامة المقام

يجب أن يجيب التقرير على السؤال بوضوح: "90 في المائة من ماذا بالضبط هو 90 في المائة؟" المقامات الممكنة:

- المشاركون المعينون
- مبادرون بالمهام
- المكملون للمهام
- السجلات المتوافقة مع البروتوكول
- نتائج النظام الصالحة
- الاستجابات مع قرارات الحكم
- خانة اللغة أو الدولة المحددة

يجب عدم تغيير المقام بعد رؤية النتيجة. على سبيل المثال:

- المهام المعينة لـ 1,000,
- المهام المكتملة لـ 900,
- المهام الصالحة لـ 800,
- المهام الناجحة لـ 720

هكذا:

```
720/800=90%
```

هذا هو معدل النجاح بين السجلات الصالحة. أما معدل النجاح الملاحظ عبر العينة المعينة بأكملها فهو:

```
720/1000=72%
```

يجيب على سؤال مختلف. يمكن استخدام كلا المعدلين. لا يمكن تقديمهما بنفس المعنى.

14. الدقة والاتساق ليستا نفس الشيء

قد يكون لتوزيع ما دقة عالية. ومع ذلك، قد يكون غير متسق مع الزمن أو بين المستخدمين. قد يكون لتوزيع آخر دقة منخفضة. ومع ذلك، قد يكون دائمًا خاطئًا. لذلك، يجب التمييز بين أربعة حالات.

الدقة	الاتساق	تعليق
عالي	عالي	مرشح للتمثيل المتسق والدقيق
عالي	منخفض	دقة غالبًا ما تكون عالية لكن تجربة غير موثوقة
منخفض	عالي	تمثيل خاطئ منهجي ومتسق
منخفض	منخفض	تمثيل غير صحيح وغير مستقر

كونه "متسقًا" ليس بالضرورة أمرًا إيجابيًا تلقائيًا. قد يكون النظام متسقًا إلى حد كبير إذا أعطى نفس الخطأ لكل مستخدم.

14.1. التناسق بين المستخدمين

هل يرى المستخدمون المختلفون تمثيلًا مشابهًا للمواد في نفس الخلية؟

14.2. التناسق بين الجلسات

هل يرى نفس المستخدم أو المستخدم المطابق نتائج مشابهة في جلسات جديدة؟

14.3. التناسق بين الموجات

هل ينتج نفس البروتوكول توزيعًا مشابهًا في أيام أو فترات مختلفة؟

14.4. اتساق ما بعد تحديث المنتج

عندما يحدث تغيير جوهري في منتج الذكاء الاصطناعي، هل يستمر التوزيع السابق؟

15. إنتاج نفس الجملة ليس استقرارًا

الاستقرار: لا يعني "تقديم نفس الإجابة لكل مستخدم كلمة بكلمة." يمكن لمنتج الذكاء الاصطناعي استخدام تعابير وترتيب وأمثلة مختلفة. طالما أنها تحافظ على المعنى الجوهري، يمكن أن تكون أشكالًا مقبولة. التمييز المهم هو:

التنوع الرسمي

اختيار الكلمات ترتيب الجملة الطول النبرة المثال استخدام قائمة أو فقرة إذا تم الحفاظ على الحقيقة الجوهرية، قد لا يكون خطأ.

التنوع الدلالي

تغيير هوية الشركة
تمييز الفئة الرئيسية
توسع أو تقلص مجال النشاط
تغيير الترخيص أو الكفاءة
تغيير السعر والجغرافيا
التوصية لمستخدم واحد دون التوصية لمستخدم مشابه بدون شرح
إذا تغير التمثيل المادي، قد تكون هناك مشكلة في الاستقرار. لذلك، المفهوم المفضل هو:

الاستقرار الدلالي المشروط

يجب أن يكون كذلك. من المتوقع الحفاظ على الواقع الأساسي في ظروف مادية مكافئة. قد يكون التغيير في التوصية بسبب احتياجات المستخدم المختلفة سلوك سياقي مناسب.

16. حساسية السياق والتغير التعسفي

ليس من الصحيح دائمًا أن يقدم النظام نفس التوصية لجميع المستخدمين. على سبيل المثال:

- المستخدمون ذوو الميزانية المحدودة
- فقط المستخدمون الذين يبحثون عن خدمات في دولة معينة
- المستخدمون الذين يحتاجون إلى رخصة مختلفة
- المستخدمون الباحثون عن خدمات فردية بدلًا من خدمات الشركات

يمكن أن يحصل على نتائج مطابقة مختلفة لنفس الكيان. هذا الاختلاف:

- قد يكون بسبب حالة المستخدم,
- نطاق الخدمة
- أو الحد الفعلي

وقد يكون حساسًا للسياق ودقيقًا. ومع ذلك، إذا تغيرت النتيجة بدون سبب في ظروف مستخدم متكافئة، فقد يكون ذلك تباينًا تعسفيًا. لذلك، لا تريد NOMOS: "نفس الإجابة لجميع المستخدمين." إنها تريد:

> ردًا يكون متسقًا وقابلًا للتفسير في ظروف مادية متكافئة وحساسًا للحدود الفعلية في ظروف مختلفة.

17. الوقت ضمن توزيع التمثيل

توزيع التمثيل ليس بلا زمن. يمكن التفكير فيه كـ PE,t. توزيع نفس الكيان:

- تحديث المنتج،
- تغيير الموارد،
- تغيير الشركة،
- حدث إخباري،
- نسخة لغة جديدة،
- تغيير الوصول التقني

يمكن أن تتباين فيما بعد.

17.1. تغيير مفاجئ

بعد تحديث منتج أو مورد، تتغير التوزيعات بسرعة في وقت قصير.

17.2. تغيير تدريجي

بسبب الموارد الجديدة وسلوكيات المستخدمين، تتغير النتيجة تدريجياً مع مرور الوقت.

17.3. تقلب مؤقت

خلال حدث محدد أو فترة أخبار، تختلف التوزيعات مؤقتاً.

17.4. تغيير بعد التدخل

بعد تدخل GEO، قد يتغير التوزيع. ومع ذلك، فإن الادعاء بأن التغيير ناتج عن التدخل يتطلب أيضًا خط أساس وتصميم السببية.

17.5. التنعيم باستخدام المتوسط الزمني

لمنتج:

- في الموجة الأولى 98,
- في الموجة الثانية 90,
- في الموجة الثالثة 82

إذا تم احتسابه، يكون المتوسط 90. ومع ذلك، فإن نشر 90 يخفي الاتجاه الهبوطي. لذلك:

> لا يمكن لمتوسط الوقت أن يحل محل نتائج الموجة.

18. مخاطر الذيل وتوزيع الأخطاء الكبيرة

قد يكون جزء كبير من التوزيع صحيحًا. ومع ذلك، يجب أيضًا تقييم الأخطاء الكبيرة المنخفضة الاحتمالية. أمثلة:

- رخصة صحية غير موجودة فعليًا
- إرشادات الطوارئ غير الصحيحة
- السلطة القانونية غير الصحيحة
- ضمان استثمار زائف
- هوية شركة غير صحيحة

- نصيحة غير مناسبة للأطفال أو الفئات الضعيفة
- تكرار التعليمات التلاعبية المخفية عن الناس

في البنية السابقة:

- حرج,
- كبير,
- متوسط,
- استشاري

تم تصنيفها إلى أربعة مستويات من الأهمية، وتم اقتراح أن النتائج الحرجة يجب أن يكون لها تأثير فشل تلقائي. يربط هذا الكتاب هذا النظام بتوزيع السكان.

18.1. فصل التكرار والشدة

خطأ:

- متكرر لكنه منخفض التأثير،
- نادراً لكنه شديد

قد يحدث. ومع ذلك، لا يمكن لمعدل التكرار وحده تقييم الحالتين بالتساوي.

18.2. تمثيل المخاطر المرجح بالشدة للمرشحين

مفهوم المرشح

يمكن تطوير مؤشر المخاطر باستخدام معاملات الشدة لفئات الأخطاء:

```
tau= Σ_{k=1}^K λ_k p_k
```

هنا:

- pk: معدل الخطأ المتوقع من النوع
- λk: معامل شدة نوع الخطأ

ومع ذلك، لا يمكن تحديد المعاملات بشكل عشوائي. يجب أن تستند إلى:

- الضرر البشري المتوقع
- سلامة التمثيل
- تأثيره على قرار المستخدم
- القابلية للإرجاع
- النية والتلاعب
- منطقة عالية المخاطر

سيتم تحديد الطريقة الدقيقة في القسم 15.

18.3. البوابة الحرجة

لا يمكن لمعدل عام مرتفع أن يحذف تلقائياً مرشح خطأ حرج تم التحقق منه. يجب أن يظهر السجل العام الصحيح كلا النتيجتين معاً: "تقدير التمثيل العام مرتفع." و: "الحدث الحرج المحدد أو الأحداث تؤثر أيضاً على قرار المطابقة."

19. الرفض والفشل في التوزيع

النظام غير المستجيب هو أيضًا جزء من تجربة المستخدم. ومع ذلك، ليست جميع حالات الرفض متشابهة.

19.1. الرفض المناسب

قد يمارس النظام الحذر عندما تكون الأدلة غير كافية أو عندما لا يستطيع تقديم استجابة آمنة. ويمكن أن يكون هذا التصرف السليم في بعض السياقات.

19.2. الرفض غير الضروري

قد يتجنب بشكل غير متسق الرد على المعلومات المتاحة علنًا والمنخفضة المخاطر.

19.3. عدم الاستجابة الفني

قد لا ينتج النظام استجابة بسبب خطأ فني.

19.4. رفض خارج النطاق

قد يكون الطلب غير صحيح أو خارج بروتوكول المهمة.

19.5. الاحتفاظ بالرفض في التوزيعات

الرفض:

- نجاح،
- فشل،

مجهول

لا ينبغي إجبارهم على الانتماء إلى فئة واحدة. يجب تقييم وظيفتهم ومبرراتهم. قد يؤدي إزالة الرفض من المقام إلى تحقيق معدل دقة عالٍ فقط بين الاستجابات الناجحة وإخفاء تجربة المستخدم الحقيقية.

20. كيف يمكن مقارنة التوزيعات؟

إن مقارنة أرقام النتائج ببساطة غير كافية لمقارنة توزيع منتجين للذكاء الاصطناعي. لكي تكون المقارنة عادلة، يجب أن تكون المجالات التالية متساوية قدر الإمكان أو مُدرجة في النموذج بشكل صريح:

- الكيان الذي تم تدقيقه
- السكان المستهدفون
- نسخة المُدخل
- اللغات
- الدول
- موجة القياس
- حالة الجلسة
- واجهة المستخدم
- توزيع الخطة
- قاعدة التحكيم
- حزمة الواقع

- قواعد الصلاحية
- الأوزان

لا يمكن استخدام نتيجتين تم الحصول عليهما باستخدام بروتوكولات مختلفة جوهريًا مباشرة في التصنيف.

20.1. مقارنة على أساس مشترك

جميع منتجات الذكاء الاصطناعي المقارنة:

- متاحة رسميًا,
- تدعم نفس اللغة وظروف المستخدم

يمكن مقارنتها في بيئة مشتركة. هذا يوفر مقارنة أنظف بين الأنظمة.

20.2. توزيع الوصول الطبيعي

يمكن قياس كل منتج ذكاء اصطناعي في عالم الوصول الفعلي لمستخدميه. تُظهر هذه النتيجة التأثير العالمي الفعلي للمنتج. ومع ذلك، نظرًا لأن الفئات المستهدفة قد تختلف، يجب توخي الحذر عند تصنيف الأداء مباشرة.

20.3. واجهات مختلفة لنفس المزود

يمكن للمنتجات على الويب، أو الهاتف المحمول، أو المنتجات المؤسسية، أو الخطط المختلفة، حتى لو كانت تابعة لنفس المزود، أن تخلق توزيعات منفصلة. إذا كان من المقرر دمجها تحت نتيجة واحدة، يجب توضيح طريقة الدمج والتوزين.

21. التوزين يغير معنى التوزيع

يعتمد المتوسط العالمي على الوزن الذي يتم به دمج الملاحظات. الأوزان الممكنة:

- متساوية حسب الدولة
- حسب عدد سكان الدولة
- عدد مستخدمي الذكاء الاصطناعي المناسب
- مستخدم اللغة
- التعرض الفعلي للمنتج
- منتج الذكاء الاصطناعي المتساوي
- حصة السوق
- السوق المستهدف للعملاء

كل طريقة تجيب على سؤال مختلف.

21.1. وزن الدولة المتساوي

يمكن أن تحصل سان مارينو والهند على نفس الوزن. يمكن أن تعالج هذه الطريقة سؤال 'تجربة الدولة المتوسطة.' لا تمثل متوسط تجربة المستخدم في سكان العالم.

21.2. وزن السكان

الدول الأكبر لها تأثير أكبر. يمكن لهذا النهج أن يقترب من تجربة السكان العالمية. قد يجعل الدول واللغات الصغيرة غير مرئية.

21.3. وزن العدالة اللغوية

يمكنه أيضًا ضمان بقاء اللغات منخفضة الموارد مرئية. ليس هو نفسه درجة السكان العالمية.

21.4. وزن تعرض المنتج

يمكن النظر في وزن منتجات الذكاء الاصطناعي وفقًا للتعرض الفعلي للمستخدمين. يتطلب ذلك بيانات استخدام موثوقة ومستقلة.

21.5. مبدأ الوزن المعياري

> يجب تحديد الأوزان قبل رؤية النتائج، والإفصاح عنها علنًا، وتحديد السؤال الذي تجيب عنه.

يُحظر اختيار نظام تحكيم يزيد النتيجة الإيجابية بعد رؤية النتيجة. بالإضافة إلى نتيجة عالمية واحدة، يمكن أيضًا نشر نتائج موزونة بديلة ذات معنى.

22. توزيع الدرجات هو ضغط فقدي

من الممكن إنتاج رقم واحد من التوزيع. ومع ذلك، فإن رقمًا واحدًا لا يحمل كل المعلومات في التوزيع. لذلك، فإن النتيجة:

هي ملخص فقدي للتوزيع

يجب قبوله. قد تفقد النتيجة المعلومات التالية:

- أضعف لغة
- أضعف دولة
- نوع الخطأ الحرج
- الكتلة غير المعروفة
- تغير الوقت
- الاستقطاب
- اختلاف المنتج والخطة
- السكان خارج النطاق
- توزيع الرفض
- غموض الحكم

النقاط مفيدة لأنها:

- تسهل المقارنة،
- تبسط التواصل العام،
- تمكن من متابعة التغيير،

تدعم القرار الإداري. ومع ذلك، لا يمكن لبساطة النقاط أن تتحول إلى إزالة التوزيع.

23. بطاقة الحد الأدنى من النتائج العامة

متطلبات المرشح — التزام المرشح

إذا كان من المقرر عرض نتيجة عالمية NOMOS للجمهور، فمن المستحسن عرض النتائج التالية على الأقل معها:

المجال	الوصف
تقدير التمثيل الرئيسي	النتيجة الإجمالية وفقًا للأوزان المحددة

المجال	الوصف
نطاق الثقة أو عدم اليقين	حساسية التنبؤ
النطاق	البلد، اللغة، المنتج، واجهة المستخدم، والوقت
أضعف مجموعة فرعية كافية	تجعل عدم المساواة المادية ظاهرًا
معدل الخطأ الحرج	قائمة الأخطاء الشديدة
استقرار الزمن	التباين بين الموجات
معدل UNKNOWN وعديم الاستنتاج	الكتلة غير المعروفة
حقول لم يتم اختبارها	الكون غير المقاس
تاريخ القياس	الحد الزمني للنتيجة
نسخة البروتوكول	أي قاعدة تم إنتاجها بها

يجب عدم استخدام النتيجة الرئيسية كعلامة إعلانية مستقلة. يجب ربط السجل الأساسي بهذه بطاقة النتائج.

24. التوزيع الاصطناعي APPLE.COM

عرض المنهجية الاصطناعية / منتجات الذكاء الاصطناعي، والمستخدمين، واللغات، والأرقام، والنتائج أدناه كلها خيالية تمامًا. إنها لا تمثل الأداء الفعلي لشركة أبل أو أي منتج ذكاء اصطناعي حقيقي. دعونا نعتبر منتجين اصطناعيين للذكاء الاصطناعي:

- المنتج أ
- المنتج ب

لكل منتج:

- خمس لغات،
- 200 ملاحظات اصطناعية صالحة لكل لغة،
- مجموع 1,000 ملاحظة مفترض.

يتم افتراض توزيع متساوٍ للغات فقط لتوضيح الطريقة. إنه لا يمثل الوزن السكاني الفعلي GEO-1000.

24.1. معدلات النجاح حسب اللغة

اللغة	المنتج أ	المنتج ب
الإنجليزية	90%	99%
التركية	91%	98%
الألمانية	89%	97%
اليابانية	90%	86%
العربية	90%	70%
المتوسط البسيط	90%	90%

متوسط المنتجين هو نفسه. ومع ذلك، توزيعاتهما مختلفة.

المنتج أ

نتائج اللغات متقاربة مع بعضها البعض. الفرق بين أعلى لغة وأدناها هو 2 نقاط. الحد الأدنى للغة هو .89%

المنتج ب

يحقق نتائج مرتفعة جدًا في ثلاث لغات. هناك تراجع في اليابانية وخاصة في العربية. الفرق بين أعلى لغة وأدناها هو 29 نقاط. الحد الأدنى للغة هو .70% ملاحظة: قول "كلاهما حصل على 90" يخفي الفرق المادي.

24.2. توزيع حالة الاستجابة

حالة الاستجابة	المنتج أ	المنتج ب
استيفاء معيار الانتقال	900	900
مدعوم جزئياً	60	20
ادعاء بمادة غير مدعومة	20	20
قديم	10	20
مرشح الخطأ الحرج	1	14
لا نتيجة / UNKNOWN	9	26
الإجمالي	1,000	1,000

معدل النجاح للمنتجين الاثنين هو مرة أخرى .90% ومع ذلك: في المنتج B، عدد المرشحين للأخطاء الحرجة أعلى بمقدار 14 مرة. في المنتج B، الكتلة المجهولة وغير الحاسمة أعلى. جزء أكبر من حالات الفشل في المنتج A يرجع إلى نقص الدعم الجزئي. تراكم الأخطاء في المنتج B أثقل. معدل النجاح نفسه لا يؤدي بالضرورة إلى نفس نتيجة الملاءمة.

24.3. موجات الوقت

موجة القياس	المنتج أ	المنتج ب
الموجة 1	90%	98%
الموجة 2	91%	90%
الموجة 3	89%	82%
المتوسط	90%	90%

المنتج أ مستقر نسبيًا. المنتج ب يظهر تراجعًا ملحوظًا. نشر متوسط الثلاث موجات فقط يخفي التدهور في المنتج ب.

24.4. نتيجة تركيبية

قد يكون من الممكن نشر نفس النتيجة الرئيسية للمنتج أ والمنتج ب:

90

لكن بطاقات نتائج NOMOS لا يمكن أن تكون نفسها.

المنتج أ — توقيع التوزيع التركيبي

النتيجة العامة: مرتفعة عدم مساواة اللغة: منخفضة ذيل الأخطاء الحرجة: منخفضة ولكن ليست صفرية الكتلة المجهولة: منخفضة استقرار الزمن: مرتفع

المنتج ب — توقيع التوزيع التركيبي

النتيجة العامة: مرتفعة عدم مساواة اللغة: مرتفعة ذيل الأخطاء الحرجة: كبير الكتلة المجهولة: أعلى استقرار الزمن: منخفض ومتراجع

لذلك، الاستنتاج الاصطناعي للقسم 2 هو كما يلي:

الحصول على نفس المتوسط للمنتج A والمنتج B لا يشير إلى أن لديهم نفس جودة التمثيل.

25. الأحكام الإلزامية المعيارية

CH02-N01

لا يمكن تعريف نتيجة GEO بواسطة متوسط واحد فقط أو معدل نجاح واحد فقط.

CH02-N02

كل نتيجة توزيع:

- الكيان الذي تم تدقيقه،
- السكان المستهدفون،
- منتج الذكاء الاصطناعي و واجهة المستخدم
- البلد واللغة،
- المُدخل,
- حالة الإقامة،
- مع الوقت

يجب أن تكون مرتبطة.

CH02-N03

المتوسط العالمي وحده لا يمكن أن يعوّض عدم المساواة في مجموعة المواد.

CH02-N04

يجب الإبلاغ عن النتائج الحرجة وأخطاء الشدة العالية الأخرى بشكل منفصل عن معدل الخطأ العام.

CH02-N05

يجب تمييز حالات UNKNOWN و no-result و rejected و invalid و NOT TESTED عن بعضها البعض.

CH02-N06

لا يمكن عرض بلد أو لغة أو منتج أو واجهة المستخدم أو فترة زمنية غير مقاسة كجزء من التوزيع الإيجابي.

CH02-N07

يجب تحديد الأوزان وقواعد التجميع قبل عرض النتائج.

CH02-N08

يجب أن تُنشئ تغييرات الوزن بعد النتائج نسخة تحليل جديدة، ويجب الحفاظ على النتائج السابقة.

CH02-N09

لا يمكن مقارنة التوزيعات المستندة إلى مجموعات مستهدفة مادية مختلفة، أو الطلب، أو المنتج، أو الوقت، أو معايير التقييم مباشرة دون شرح.

CH02-N10

لا يمكن اعتبار التغير في اللغة الرسمية خطأً في التمثيل تلقائيًا.

CH02-N11

لا يمكن تصنيف التباين في هوية المادة، أو النطاق، أو الوقت، أو الكفاءة، أو التوصية كاختلاف في الأسلوب دون تبرير.

CH02-N12

الاستقرار العالي وحده ليس دليلًا على الجودة؛ يجب أيضًا تسجيل النتائج الخطأ المتكررة.

CH02-N13

يجب أن تحتوي النتائج العامة على قيود التغطية إذا كان متوسط مرتفع مصحوبًا بتغطية منخفضة.

CH02-N14

لا ينبغي نشر نتائج المجموعات الفرعية كأحكام نهائية على بلد أو لغة في ظل عينة غير كافية وعدم اليقين الكبير.

CH02-N15

يجب الحفاظ على معدلات الرفض والفشل الفني كجزء من تجربة المستخدم الحقيقية.

CH02-N16

يجب أن تكون النتيجة الوحيدة المنشورة مرتبطة بالتوزيع القانوني وسجل النطاق.

CH02-N17

لا يمكن دمج التوزيعات الاصطناعية مع أداء منتج ذكاء اصطناعي حقيقي أو كيان حقيقي.

CH02-N18

يجب أن يحتوي تقرير التوزيع على مالك بشري أو مؤسسي مسؤول.

26. أشكال الفشل

CH02-F01 — التفكير المتوسط

يتم عرض المتوسط العالمي العالي؛ يتم تخزين اللغة أو البلد أو خطأ فادح في الفساد.

CH02-F02 — حذف المجموعة الفرعية

يتم إزالة المجموعات منخفضة الأداء من التقرير أو الإعلان عنها على أنها 'غير مهمة'.

CH02-F03 — خريطة النطاق

تُستخدم النتائج الجيدة ذات النطاق الضيق لعالم بأسره، جميع اللغات، أو جميع المنتجات.

CH02-F04 — توحيد شدة الخطأ

يتم دمج النواقص الطفيفة والأخطاء الصحية الخطيرة أو المتعلقة بالترخيص ضمن عدد الأخطاء نفسه.

CH02-F05 — إخفاء التدهور باستخدام المتوسط الزمني

يتم جعل التراجع أو التذبذب غير مرئي فقط باستخدام متوسط الفترة.

CH02-F06 — عبادة الاستقرار

يُفترض أن الإجابة المتسقة صحيحة تلقائيًا.

CH02-F07— مغالطة الإجابة المتماثلة كلمة بكلمة

يُستخدم مطابقة النص الحرفي كمقياس للاتساق بدلاً من الصحة الدلالية.

CH02-F08— احتساب الاختلاف السياقي كخطأ

يتم تمييز الاختلافات في التوصيات المناسبة بناءً على احتياجات المستخدم الفعلية على أنها تقلبات تعسفية.

CH02-F09— احتساب التقلبات التعسفية كتحسين شخصي

يتم الدفاع عن التناقضات المادية في ظروف المستخدم المتساوية على أنها "تخصيص" دون تفسير.

CH02-F10— تغيير المقام

يتم إجراء الانتقال بعد النتيجة بين المقامات الملاحظة المكتملة أو الصالحة أو المقيمة لرفع النتيجة الإيجابية.

CH02-F11— الترجيح بعد النتيجة

يتم بعد ذلك اختيار الدولة أو اللغة أو معاملات النموذج التي تنتج أعلى نتيجة.

CH02-F12— دمج الخلايا غير المتسقة

يتم عرض شروط مختلفة من المُدخل أو المنتج أو الوقت أو اللوحة كتوزيع واحد.

CH02-F13— حذف الرفض

يتم إزالة سجلات النظام التي لم تستجب لتوليد دقة عالية فقط بين الذين قدموا استجابة.

CH02-F14— الحكم المؤكد من الخلية الصغيرة

يتم إنتاج درجة البلد أو اللغة من مستخدم واحد أو عدة مستخدمين.

CH02-F15— إخفاء الذيل الحرج

يتم فقدان الأخطاء الشديدة النادرة ضمن معدل النجاح الكلي.

CH02-F16— تقرير درجة واحدة

يتم نشر الدرجة الرئيسية فقط دون إظهار النطاق، وعدم المساواة، والاستقرار، وعدم اليقين، والأخطاء الشديدة.

CH02-F17— عد التوزيع على أنه "اعتقاد الذكاء الاصطناعي"

يتم تقديم نتائج واجهة المستخدم الملاحظة كما لو كانت وجهة نظر ثابتة وجوهرية للذكاء الاصطناعي.

CH02-F18— عد غير المقيس على أنه صفر التباين

تُعتبر المناطق غير المختبرة سلسة ومستقرة.

27. إجراء التدقيق

ينبغي فحص ما إذا كان تقرير GEO يمثل التوزيع بصدق من خلال الخطوات التالية.

الخطوة 1— تسجيل الادعاء العام

مثال: "العلامة التجارية حصلت على درجة NOMOS العالمية وهي 92." يتم تسجيل الجملة الكاملة والواجهة الذي وُجدت عليه.

الخطوة 2— تحديد التوزيع المستهدف

لأي درجة:

- مجموعة المستخدمين،
- منتجات الذكاء الاصطناعي،

- الدول،
- اللغات،
- الفترة الزمنية

يتم تطبيقها؟

الخطوة —3 استخراج خلايا القياس

يتم تحديد كل تركيبة مادة:

- المنتج
- البلد
- اللغة
- العينة
- مُدخل
- الموجة
- واجهة المستخدم

الخطوة —4 إعادة بناء المقامات

يجب الإبلاغ عن المهام المعينة، والمهام المكتملة، والسجلات الصالحة، والسجلات المحكَّمة، والاستجابات الناجحة كأعداد منفصلة.

الخطوة —5 استخراج حالات النتائج

ليس فقط معدل النجاح:

- جزئي،
- غير مدعوم،
- متعارض،
- قديم،
- حرج,

غير معروف،

مرفوض، أو غير حاسم يتم فحص التوزيع.

الخطوة —6 التحقق من الأوزان

الأوزان:

- هل هي محددة مسبقًا,
- أي سؤال يجيبون عليه،

هل تغيروا بعد النتيجة؟

الخطوة —7 فحص اختلافات الفئات الفرعية

يتم تقييم أفضل مجموعة، أضعف مجموعة كافية، اختلاف اللغة، اختلاف البلد، اختلاف المنتج، واختلاف اللوحة.

الخطوة 8 — مراجعة النطاق والنقص

ما هي المجالات التي لم يتم قياسها؟ أي الخلايا غير كافية؟ ما هي أسباب النقص؟ هل يمكن أن تكون البيانات المفقودة منهجية؟

الخطوة 9 — فحص التحديد

يتم تقييم التغيير بين جلسات موجات المستخدمين والواجهات.

الخطوة 10 — فحص خطر الطابور

الأخطاء الحرجة وغيرها من الأخطاء الجسيمة:

- الرقم،
- معدلها،
- النوع،
- تأثير المستخدم

يتم تقييمه بشكل منفصل من حيث

الخطوة 11 — قارن الملخص العام مع التوزيع

هل الرقم أو الجملة الفردية المنشورة لعامة الناس تعكس بدقة اتجاه ومخاطر التوزيع الكامل؟

الخطوة 12 — حدد النتيجة

يدعي الإعلان العام فقط ما يحمله التوزيع والنطاق.

28. الدليل الضروري

إلى الحد الذي يكون ذا صلة بتوزيع ممثل أو درجة جي.إي.أو العالمية، يجب البحث في السجلات التالية:

- سجل الكيان المدقق
- تعريف السكان المستهدفين
- إطار العينة
- منتج الذكاء الاصطناعي وواجهات المستخدم
- خلايا الدولة واللغة
- الإرشادات والإصدارات
- ظروف الجلسة واللوحة
- موجات القياس
- أرقام المهام المعينة
- المهام المكتملة
- السجلات الصحيحة وغير الصحيحة
- الردود الخام
- توزيع حالة النتائج
- سجلات التحكيم
- سجلات الأخطاء الحرجة
- الرفض والسجلات غير الحاسمة

- سجلات UNKNOWN
- الأوزان
- أصحاب المصلحة
- نتائج المجموعة الفرعية
- نتائج الاستقرار
- المجالات خارج النطاق
- حسابات الثقة أو عدم اليقين
- بطاقة الأداء العامة
- الشخص أو المؤسسة المسؤولة
- إصدار البروتوكول والتحليل
- سجل التغييرات

29. قائمة تدقيق التدقيق

هل الفئة المستهدفة للتوزيع واضحة؟ هل منتجات الذكاء الاصطناعي المدققة وأسواق المستخدم محددة؟ هل البلاد واللغة والنطاق الزمني والمُدخلات مرئية؟ هل تم الكشف عن الأوزان المستخدمة لإنتاج النتيجة المركبة؟ هل المقام معرف بالكامل؟ هل يتم الإبلاغ عن السجلات غير الصالحة بشكل منفصل؟ هل يتم الاحتفاظ بالرفض والنتائج غير الحاسمة؟ هل UNKNOWN و NOT TESTED متميزتان؟ هل نتائج المجموعات الفرعية مرئية؟ هل تم تفسير الخلايا الصغيرة بثقة مفرطة؟ هل تم الاحتفاظ بأضعف مجموعة مادة؟ هل يتم الإبلاغ عن الأخطاء الحرجة بشكل منفصل؟ هل يتم عرض موجات القياس بشكل منفصل؟ هل المتوسط يخفي التراجع أو التقلب؟ هل يتم تمييز الفرق الشكلي فقط عن الفرق المادي في المعنى؟ هل تم تصنيف التباين الحساس للسياق في التوصيات بشكل صحيح؟ هل تم دمج الواجهات المختلفة من نفس المزود دون تفسير؟

هل تم تضمين المناطق خارج النطاق في النتيجة العالمية؟ هل تم فصل التوزيعات الاصطناعية والحقيقية؟ هل تعكس الجملة المقدمة للجمهور الاتجاه الحقيقي للتوزيع الكامل؟ إذا لم يتم الرد على أحد الأسئلة الجوهرية، فإن نتيجة التوزيع تحمل ثقة محدودة. إذا لم يتم الرد على أكثر من سؤال واحد، فلا ينبغي تقديم درجة عامة واحدة كحكم نهائي للجمهور.

30. الاعتراضات والإجابات

الاعتراض 1 — "يريد الجمهور رقمًا واحدًا."

هذا صحيح. يمكن لرقم واحد أن يسهل عملية التواصل. NOMOS لا يرفض الرقم الواحد بشكل كامل. ومع ذلك، فإن الرقم الواحد:

- النطاق،
- خطأ حرج
- قاعدة المجموعة الفرعية،
- الوقت،
- عدم اليقين

يجب أن تكون مرتبطة بالمعلومات. يمكن أن يكون الرقم قصيرًا. لا يمكنه تقصير الحقيقة.

اعتراض —2 'اللغات الصغيرة تقلل من النقاط العالمية بلا داعٍ.'

يمكن أن يأخذ وزن السكان العالمي في الاعتبار لغة صغيرة بوزن محدود. ومع ذلك، لا يمكن تجاهل التحريف الكبير في التمثيل في لغة صغيرة لمجرد أن حصتها السكانية منخفضة. لذلك:

- الدرجة الرئيسية الموزونة بالسكان،
- كنتيجة للعدالة اللغوية المنفصلة

يمكن نشرها معًا. المقياسان يجيبان عن أسئلة مختلفة.

الاعتراض —3 'إذا كان هناك تخصيص، فإن الاختلاف في الردود أمر طبيعي.'

بعض الاختلافات يمكن أن تكون طبيعية وصحيحة. ومع ذلك، فإن التخصيص:

- هوية شركة،
- رخصة،
- سعره الحالي،
- منطقة الخدمة الحقيقية

لا ينبغي تغييره بشكل عشوائي. يمكن أن يؤثر التخصيص على التفسير والنصيحة المناسبة للمستخدم. فهو ليس إذنًا لاختراع حقيقة مادية.

الاعتراض —4 'لماذا سيكون مشكلة إذا لم يتم إعطاء نفس الإجابة؟'

المشكلة ليست في اختلاف الكلمات. المشكلة هي تغيير الواقع المادي. NOMOS لا تبحث عن المطابقة النصية الدقيقة، بل عن التناسق الدلالي الشرطي.

الاعتراض —5 "لماذا سيكون مهمًا إذا حدث خطأ حرج مرة واحدة فقط في إجابات 1,000؟"

قد يكون هذا الخطأ نادرًا من حيث انتشار السكان. ومع ذلك، قد يكون شديد الضرر. فقدان نظام حرج على متن الطائرة مرة واحدة في كل ألف رحلة لا يُقييم فقط كنسبة 0.1 في المئة. في سياق GEO، يجب أيضًا فحص التكرار والضرر بشكل منفصل.

الاعتراض —6 "لا يمكننا العثور على عينة كافية لكل دولة صغيرة أو لغة."

هذا صحيح. لذلك:

- لوحة السكان العالمية,
- لوحة مراقبي الدول,
- لوحة عدالة اللغة

يمكن تأسيسه بشكل منفصل. الملاحظة الفردية لا تتحول إلى درجة الدولة. فهي تشير فقط إلى أن هناك ملاحظة موجودة. الحالة الصحيحة لعينة غير كافية:

عينة غير كافية

أو:

لم يتم تقديرها

يجب أن تكون.

اعتراض —7 "هذه التفاصيل كثيرة ومعقدة جدًا ليفهمها العميل."

ليس من الضروري أن يرى العميل جميع الصيغ على الصفحة الرئيسية. ومع ذلك، لا يمكن إخفاء الحقائق الأساسية اللازمة لاتخاذ القرار. يجب أن تجيب بطاقة النتائج المختصرة على الأسئلة التالية: ما هي النتيجة العامة؟ ما هو

الخطر الأكثر خطورة؟ ما هو المجال الأضعف؟ ما الذي لم يتم اختباره؟ ما مدى قوة النتيجة؟ إلى أي تاريخ تنتمي؟ يجب إدارة التعقيد، وليس التخلص منه.

اعتراض 8— "إذا كان نفس الطراز يُستخدم في جميع البلدان، فلماذا يجب أن يكون هناك اختلاف بين البلدان؟"

المخرجات التي يراها المستخدم قد لا تتكون فقط من النموذج وحده. تشمل الاختلافات بين البلدان ما يلي:

- الوصول إلى المنتج،
- الموارد الإلكترونية,
- المكانية,
- اللغة،
- الخطة،
- واجهة
- التشريع،
- التخصيص

يمكن أن يظهر من خلاله. قد لا يحدث فرق. ومع ذلك، لا يمكن افتراض غيابه دون قياس.

اعتراض 9— "التسجيل الموثوق به غير ممكن لأن استجابات الذكاء الاصطناعي متغيرة."

التغيرية لا تجعل القياس مستحيلاً. على العكس، إنها إحدى الخصائص التي تحتاج إلى القياس. النتيجة:

- تاريخ محدد،
- النطاق،
- فاصل الثقة،
- معلومات عن الاستقرار

قد تكون ذات معنى طالما تحملها. ومع ذلك، فإن الادعاء باليقين الأبدي لا يمكن الدفاع عنه.

اعتراض 10— "إذا تغير الطلب، يتغير التوزيع بالكامل."

هذا صحيح. لذلك، التوزيع مشروط بالمُدخل. GEO-1000 لا ينتج الإجابة الواحدة لكل الأسئلة الممكنة. معرف مسبقًا:

- مُدخل المرآة الأساسية،
- سجل المُدخل التشخيصي

يقيس مجالات تمثيلية محددة. سيتم تحديد دستور المُدخل في القسم 10.

القاعدة العامة للقسم 32

لم يتم كتابة هذا القسم ضد درجة واحدة فقط. بل تم كتابته ضد الافتراض بأن درجة واحدة فقط تمثل كامل الواقع. الناس يريدون الأرقام. المؤسسات تريد التصنيفات. مجالس الإدارة تريد رؤية خط واحد. قد يسأل السوق: "ما هي درجتك في NOMOS؟" هذا السؤال مشروع. ولكن لا يمكن أن يكون الجواب فقط: ".90" لأنه وراء 90، قد توجد أي من الحقائق التالية:

- دقة متسقة عبر جميع اللغات
- 99 في بعض اللغات، 45 في لغات أخرى
- متوسط مرتفع ولكن خطأ حرِج شديد
- دقة عالية ولكن تغطية منخفضة

- نتائج فورية عالية ولكن تراجع سريع
- نجاح فقط في صياغات المُدخل السهلة
- معدل مرتفع فقط بين الإجابات القابلة للتقييم
- افتراض الدول غير المختبرة بشكل إيجابي
- دمج منتجات الذكاء الاصطناعي المحددة

تتشكل قيمة النتيجة من وضوح التوزيع الأساسي. إذا لم يكن التوزيع مرئيًا، فإن النتيجة لا تولد الثقة. إنها تنتج فقط مظهرًا من اليقين. وبالتالي، فإن قانون القياس الثاني لـ NOMOS هو كما يلي:

> يمكن أن يكون تسجيل *GEO* ملخصًا للتوزيع. / لا يمكنه استبدال التوزيع.

قانونه الثالث هو:

> المتوسط المرتفع لا يلغي مسألة من يتم تمثيله بشكل ضعيف.

القانون الرابع هو كما يلي:

> الخطأ الجسيم النادر ليس تافهًا لمجرد أنه نادر.

القانون الخامس هو كما يلي:

> المجال غير المختبَر ليس الجانب الإيجابي للتوزيع؛ إنه الحد المجهول.

الطلب 2 من NOMOS

> لا تَعرض لي المتوسط فقط.
>
> أظهر من هو دون المتوسط.
>
> لا تُخفي أسوأ لغتك بينما تُظهر أفضل لغتك.
>
> لا تستخدم وزن الدول الكبيرة لمحو وجود الدول الصغيرة.
>
> لا ترسل شخصًا واحدًا إلى دولة صغيرة ثم تعلن عن نتيجة نهائية لتلك الدولة.
>
> لا تقلل من شأن خطأ جسيم واحد لمجرد أنه مخفي داخل استجابة صحيحة بنسبة *99* في المائة.
>
> لا تعلن عن خطأ لمجرد أن صياغة الإجابة قد تغيرت. / ولكن عندما تتغير الحقيقة الأساسية، لا ترفضها على أنها مجرد فرق أسلوبي.
>
> لا ترمِ المرفوضات والمجهولات وغير المختبرة خارج التوزيع.
>
> لا تغير الوزن أو الحصة أو النطاق بعد رؤية النتيجة.
>
> قم بإنشاء رقم. / لكن لا تجعل الدول واللغات والمستخدمين والأوقات والأخطاء وراء الرقم غير مرئية.
>
> لأن تمثيل الكيان ليس جملة متوسطة. / إنه توزيع للحقائق المختلفة التي يواجهها أشخاص مختلفون.

جملة ختام الفصل

> موضوع قياس *GEO* ليس إجابة واحدة أو درجة متوسطة واحدة للذكاء الاصطناعي؛ بل هو مجموع النتائج التمثيلية التي يتم توليدها في مجموعة مستخدمين محددة في ظروف محددة.

النواة المعيارية

يجب التعامل مع تمثيل الكيان في الأنظمة التوليدية بوصفه توزيعًا شرطيًا للنتائج القابلة للرصد عبر أبعاد محددة تشمل: - منتجات الذكاء الاصطناعي وواجهات الاستخدام، - المستخدمين أو وحدات المعاينة، - البلدان، - اللغات والإعدادات المحلية، - حالات الجلسة والتخصيص، - المُدخلات، - وأوقات القياس. ولا يجوز لدرجة إجمالية واحدة أن تحل محل الإفصاح عن: - النطاق والتغطية، - توزيع النتائج، - التباين بين المجموعات الفرعية، - معدل الأخطاء الحرجة، - الكتلة المجهولة وغير المختبرة، - الاستقرار الزمني، - قواعد الترجيح، - وعدم اليقين. ولا يجوز تفسير تساوي المتوسطات على أنه تكافؤ في جودة التمثيل إذا اختلفت توزيعات المجموعات الفرعية أو مستويات الشدة أو التغطية أو الاستقرار اختلافًا جوهريًا. ويجب ألا يُفترض امتثال المجموعات غير المختبرة، كما يجب أن تظل النتائج النادرة ذات الضرر المرتفع ظاهرة بصورة منفصلة عن متوسط الأداء.

الفصل 03

ما الكيان الخاضع للتدقيق تحديداً؟

الأسس المنهجية / K10; K11

حد الفصل

أقرت المادة 1 الحكم التالي:

> الاستجابة الواحدة ليست درجة *GEO*.

القسم 2 أضاف النص التالي:

> تمثيل الكيان في الأنظمة التوليدية هو توزيع يحدث بين المستخدمين ومنتجات الذكاء الاصطناعي والدول واللغات والجلسات والأوقات.

يجيب هذا القسم على السؤال الإلزامي التالي:

> تمثيل من أو ماذا يقيس هذا التوزيع بالضبط؟

شركة؟ علامة تجارية؟ اسم نطاق؟ منتج؟ خدمة؟ مؤسس؟ مجموعة شركات؟ كيان قانوني محلي في دولة معينة؟ عمل امتياز؟ عندما يرى نظام الذكاء الاصطناعي نطاق apple.com فقط:

- اسم النطاق،
- العلامة التجارية,
- الشركة الأم،
- الكيان المحلي للمبيعات،
- عائلة المنتج،
- خدمة المتجر

هل سيعامل النظام جميعها ككيان واحد؟ ما لم يتم التمييز قبل بدء التدقيق، فقد يتم حساب النتيجة بشكل صحيح ولكن تطبيقها على الكائن الخاطئ. هذا الفصل:

- الكيان الذي تم تدقيقه،
- مرساة الهوية المستخدمة في المُدخل،
- الكيان الذي يحلله النظام،
- الوحدة التي سيتم تطبيق قرار المطابقة عليها،
- العلاقات بين الكيانات،
- أيه كيان تنتمي إليه الخصائص،
- حدود الهوية الزمنية والجغرافية

يحدد. هذا الفصل لم يتطرق بعد إلى:

- كيفية التحقق من جميع الادعاءات المتعلقة بالكيان،
- كيف سيقوم المحكّمون بتقييم الادعاءات الذرية،

- حصص العينات بحسب الدولة واللغة،
- صيغة النتيجة النهائية NOMOS

لم يتم الانتهاء منها بعد. مهمة القسم 3 أكثر أساسية:

> تثبيت هوية وحدود الشيء الذي سيتم قياسه قبل بدء القياس.

التحدي NOMOS

أنت تسألني هذا السؤال: "ما نوع الشركة Apple.com؟" يمكن أن يكون هناك على الأقل ثلاثة أشياء مختلفة في السؤال:

- اسم النطاق apple.com
- علامة Apple التجارية
- الشركة القانونية المرتبطة باسم النطاق والعلامة التجارية

سأختار واحدة من هذه. ربما الصحيحة. ربما الخاطئة. ربما أدمج أكثر من واحدة في إجابة واحدة. ثم ستقوم بتقييم الإجابة. لكنك لم تخبرني مسبقًا أي كيان تقوم بتقييمه. قد أكون:

- وصفت النطاق كمتجر إلكتروني،
- وصفت العلامة التجارية كعلامة تجارية للإلكترونيات الاستهلاكية،
- وصفت الشركة القانونية كشركة عالمية،

قد أكون قد أدرجت المنتجات كما لو أن الشركة نفسها هي التي تقدمها. قد تبدو إجابتي سلسة. ولكن إذا كان موضوع التدقيق غير واضح، فإن ما إذا كانت صحيحة أم لا يبقى أيضًا غير مؤكد. الآن، اعتبر مثالًا أكثر تعقيدًا.

تُدار العلامة التجارية بواسطة الشركة الأم في بلد واحد؛ وعن طريق فرع تابع في بلد آخر؛ وعن طريق موزع مستقل في بلد ثالث. يسأل أحد المستخدمين: "هل تقدم هذه العلامة التجارية خدمات في تركيا؟" أي كيان يجب أن يجيب؟ الموقع العالمي للعلامة التجارية؟ الشركة المحلية في تركيا؟ أو الموزع؟

مشغل الفرنشايز؟ هل تنتقل الرخصة المملوكة للشركة العالمية تلقائيًا إلى الفرنشايز المحلي؟ هل تثبت تقييمات العملاء على العملية المحلية جودة العلامة التجارية العالمية؟ هل يمكن استخدام عدد موظفي الشركة الأم كقدرة للشركة المحلية؟

هل يمكن نقل جائزة منتج إلى جميع الشركات التي تبيعه؟ هل تصبح خبرة المؤسس الشخصية عمر الشركة؟ هل يُعتبر مالك النطاق السابق والمالك الجديد لنفس الكيان؟

عندما تستحوذ شركة على شركة أخرى، هل تصبح جميع إنجازات الشركة السابقة سمة للشركة الجديدة؟ يمكن توليد درجة GEO دون الإجابة على هذه الأسئلة. ومع ذلك، لا يُعرف ما تمثله الدرجة. الحكم الأول في هذا القسم كما يلي:

> لا يمكن قياس الكيان حتى يتم حله.

ينص حكمه الثاني على:

> مشاركة نفس الاسم لا يعني أن يكون كيانًا واحدًا.

تنص الفقرة الثالثة منه على:

> العلاقة لا تسمح تلقائيًا بنقل الممتلكات بين الأطراف.

1. غرض الفصل

الغرض من هذا القسم هو تحديد هوية الكيان وحدوده وعلاقاته ضمن GEO-1000 قبل رؤية النتائج. يقوم القسم بالتفرقة التالية كمعيار:

- اسم النطاق والشركة القانونية
- العلامة التجارية ومالك العلامة التجارية
- الشركة الأم والشركة التابعة
- المجموعة وشركة المجموعة
- المنتج والمصنع
- الخدمة ومقدم الخدمة
- المؤسس والمؤسسة
- الموظف والمؤسسة
- حامل الترخيص والطرف الذي يستخدم العلامة التجارية
- المانح والعميل الفرعي
- المُصنّع والمُوزّع
- المنصة والبائع على المنصة
- مزوّد التكنولوجيا ومستخدم التكنولوجيا
- العميل وشريك الأعمال
- العلاقة النشطة والعلاقة التاريخية
- الاسم الرسمي والترجمة الوصفية
- كيانات مختلفة بنفس الاسم
- الإصدارات السابقة واللاحقة للكيان
- الكائن الذي تم تدقيقه والكائن الذي سيتم تطبيق قرار المطابقة عليه

في نهاية هذا القسم، يجب أن يكون كل تدقيق GEO قادرًا على تقديم إجابة واحدة وقابلة للتتبع للسؤال التالي:

> أي كيان بالضبط، مع أي علاقات، لأي فترة زمنية، وعلى أي واجهات رقمية ينتمي هذا التقييم والقرار؟

2. الأحكام المعيارية المركزية

> قبل البدء في جمع البيانات، يجب أن يحدد كل تدقيق *GEO-1000* كيانًا مراجَعًا فريدًا، وسجل كيان أساسي، وحدود نطاق، وحدود زمنية، ومخطط علاقات.

لا يكفي أي مما يلي بمفرده لتحديد كيان:

- الاسم،
- الشعار،
- اسم المجال،
- حساب وسائل التواصل الاجتماعي

هذه المعرفات وحدها لا تحدد كائن التدقيق. يجب وصف الكيان الخاضع للتدقيق، عند الضرورة، من خلال العناصر التالية على الأقل:

- نوع الكيان

- الاسم القانوني الموحد
- الاسم القانوني
- الأسماء التجارية
- الأسماء الرسمية المحلية
- الاختصارات والألقاب المستعارة
- تاريخ التأسيس أو البداية
- الاختصاص القضائي
- رقم التسجيل أو هوية رسمية معادلة
- علاقات الشركة الأم والشركة الفرعية
- ملكية العلامة التجارية
- التحكم في اسم النطاق
- ملكية المنتج والخدمة
- مسؤول تسليم الخدمة الفعلي
- المدعي
- حامل الأدلة
- الدول والمناطق النشطة
- فترة الصلاحية
- الاسم والعلاقات التاريخية
- الواجهات المدرجة في التدقيق
- الكيانات والواجهات خارج النطاق

يجب عدم نشر النتيجة الجغرافية الرئيسية حتى يتم استكمال هذا السجل.

3. ما هي الكيان؟

في هذا البروتوكول، الكيان هو كائن يمكن تخصيص هوية منفصلة وقابلة للتتبع له، أو سمة، أو علاقة، أو مسؤولية، أو حالة زمنية. الكيان:

- إنسان،
- كيان قانوني،
- مؤسسة عامة،
- علامة تجارية،
- المنتج,
- الخدمة،
- اسم المجال،
- تطبيق رقمي،
- مجموعة شركات،
- فرع مادي،
- برنامج،
- شهادة،

- منشور،
- مشروع بحثي

هذا ممكن. ليس كل عنصر مذكور يجب أن يكون كيانًا منفصلًا. ومع ذلك، إذا تغير أحد المجالات التالية بشكل جوهري، فقد يكون من الضروري تسجيل كيان منفصل:

- المسؤولية القانونية
- الملكية
- الرخصة
- تسليم المنتج أو الخدمة
- الاختصاص القضائي
- الفترة الزمنية
- الالتزام تجاه المستخدم
- الأدلة وحقوق الادعاء
- نطاق التدقيق

4. ستة عناصر منفصلة في تدقيق GEO

في تدقيق GEO، غالبًا ما تخفي كلمة "كيان" ستة وظائف منفصلة. يجب تمييز هذه العناصر الستة عن بعضها البعض.

4.1. مرساة صياغة المُدخل

مرساة صياغة المُدخل هي تعبير أو علامة تُستخدم لاستدعاء الوجود في استعلام المستخدم. أمثلة:

- اسم الشركة
- اسم العلامة التجارية
- اسم النطاق
- اسم المؤسس
- اسم المنتج
- اسم التطبيق
- اسم الشركة المحلية
- الاختصار

مثال: "ما نوع الشركة Apple.com؟" مرساة صياغة المُدخل هنا: apple.com هي اسم النطاق. لا يجب أن تكون مرساة صياغة المُدخل الكيان نفسه. فهي مجرد نقطة دخول يستخدمها النظام لتحديد الكيان.

4.2. الكيان المستهدف

الكيان المستهدف هو الكيان الذي يرغب البروتوكول فعليًا في قياسه. مثال: يمكن استهداف الكيان المؤسسي الرسمي المرتبط باسم النطاق apple.com. ومع ذلك، يجب تسجيل هويته الدقيقة قبل المراجعة. لا يمكن اختيار الكيان المستهدف من خلال النظر إلى النتائج بعد الاستفسار.

4.3. الكيان الذي حُلّ بواسطة النظام

الكيان المحلّل هو كيف يفسر منتج الذكاء الاصطناعي مرساة الاستفسار ككيان. التمثيل:

```
E_i
```

يمكن استخدامه. الكيان المستهدف:

E*

هو تقييم حل الهوية الذي يطرح السؤال الأساسي:

```
E_i= E*?
```

ومع ذلك، فإن المساواة ليست دائمًا مجرد مساواة في الاسم. يمكن للنظام التعرف على العلامة التجارية والكيان المؤسسي الصحيحين دون تسمية الكيان القانوني الصحيح. لذلك، يمكن تصنيف مطابقة الهوية لاحقًا على النحو التالي:

- مطابقة دقيقة،
- مطابقة مقبولة للعلامة التجارية،
- مطابقة جزئية،
- الكيان غير الصحيح
- غير مؤكد،
- اندماج كيانات متعددة

يمكن تصنيفه كالتالي.

4.4. موضوع الادعاء

موضوع الادعاء هو الكيان الذي ينتمي إليه ميزه محددة في الاستجابة. ضمن نفس الاستجابة:

- تاريخ التأسيس للشركة الأم،
- ميزة المنتج للمنتج،
- تقييم العميل لمزود الخدمة المحلي،
- رخصة لكيان قانوني محدد،
- خبرة المؤسس لشخص

قد ينتمي. حتى إذا بدأ الرد بالكيان المستهدف الصحيح، فقد ينقل الميزات إلى الكائنات الخاطئة. لذلك، يجب تقييم كل ادعاء واقعي بالهيكل التالي:

```
(الموضوع، العلاقة أو السمة، القيمة، الوقت، النطاق)
```

مثال: "لدى X ترخيص Y." هنا، يجب تحديد لمن:

- كيان قانوني،
- البلد،
- خلال أي تواريخ،
- للنشاط

الذي ينتمي إليه.

4.5. كائن التدقيق

كائن التدقيق هو كيان محدد أو حزمة من الكيانات حيث يتم فحص الأدلة والمحتوى والواجهة الفني ومخرجات الذكاء الاصطناعي معًا. يمكن أن يكون كائن التدقيق كيانًا واحدًا. مثال: شركة قانونية محددة أو يمكن أن يكون حزمة ذات حدود واضحة: العلامة التجارية الرئيسية + اسم النطاق القانوني + منتجين محددين + السطوح العامة باللغتين التركية والإنجليزية. إذا تم استخدام الحزمة، يجب أن تحتفظ الكيانات ضمنها بهوياتها المنفصلة. "الحزمة" لا تعني أن جميع السمات تنتقل إلى بعضها البعض.

4.6. وحدة الامتثال

وحدة الامتثال هي الكائن المحدد الذي سيتم تطبيق قرار التدقيق عليه قانونيًا ونظاميًا. قرار الامتثال:

- ينطبق على المجموعة الكاملة من الشركات،
- ينطبق على جميع اللغات،
- ينطبق على جميع أسماء النطاقات،
- على جميع المنتجات

لا يمكن تطبيقه تلقائيًا. مثال: "تم العثور على الصفحات العامة التركية والإنجليزية للعلامة التجارية NobleJackal على example.com متوافقة وفقًا للمعيار المحدد وضمن نطاق التواريخ المحدد." هذا القرار:

- جميع العمليات الداخلية للشركة،
- الشركات التابعة،
- التطبيقات المنفصلة،
- اللغات التي ستضاف في المستقبل،
- المنتجات خارج النطاق

لا يغطي تلقائيًا.

5. أنواع الكيانات

يجب تصنيف كل كيان تدقيقي بنوع واحد أو أكثر. يمكن التمييز بين النوع الأساسي والأنواع الثانوية.

5.1. الكيان القانوني

هي شركة تحمل حقوقًا ومسؤوليات منفصلة، أو تسجيلًا بموجب القانون أو اللوائح القانونية ذات الصلة. مجالاتها بقدر ما هي ذات صلة:

- الاسم القانوني
- نوع الشركة
- مكان التسجيل
- رقم التسجيل
- الهوية الضريبية أو ما يعادلها
- تاريخ التأسيس
- الحالة النشطة
- الممثل المفوض
- العلاقات بين الشركة الأم والشركات التابعة

قد يحمل الكيان القانوني نفس اسم العلامة التجارية. ومع ذلك، فهو ليس نفس الكائن.

5.2. العلامة التجارية

هو الاسم المستخدم في السوق لمنتج أو خدمة أو هوية مؤسسية. العلامة التجارية:

- قد لا تمتلك كيانًا قانونيًا منفصلًا،
- قد تكون مملوكة لشركة أخرى،
- قد تُستخدم بموجب ترخيص،

قد تُدار بواسطة شركات متعددة في دول مختلفة. يجب أيضًا تحديد الجهة المسؤولة فعليًا عن تقديم العمل الذي تنسبه العلامة التجارية.

5.3. مجموعة شركات

إنها هيكل يتكون من عدة كيانات قانونية لها ملكية أو علاقات تحكم مشتركة. لا يمكن لـ "المجموعة" نقل الميزات التالية تلقائيًا إلى جميع الأعضاء:

- الرخصة
- الشهادة
- عدد العملاء
- عدد الموظفين
- الإيرادات
- المكتب
- نتيجة التدقيق
- حالة الأمان
- علامة المطابقة

إذا كانت قيمة المستوى الجماعي موحدة حقًا، يجب ذكر ذلك بوضوح.

5.4. شركة تابعة أو شقيقة

كيان قانوني يخضع لملكية أو علاقة سيطرة لكيان آخر. الشركة التابعة:

- قد يكون طرفًا في عقد منفصل،
- حامل رخصة منفصل،
- صاحب عمل منفصل
- متصرّف بيانات مستقل

ربما. خصائص الشركة الأم لا تُنقل افتراضيًا إلى الشركة التابعة.

5.5. اسم النطاق

إنه عنوان رقمي وواجهة نشر. اسم النطاق:

- ليست كيانًا قانونيًا,
- لا يقدم الخدمات بمفرده,
- ليس من الضروري امتلاك علامة تجارية،

يمكن التحكم فيه بواسطة مالكين أو مشغلين مختلفين في أوقات معينة. يمكن أن يكون اسم النطاق نقطة البداية للتدقيق. ومع ذلك، يجب أيضًا توضيح تحت أي كيان تم اتخاذ القرار.

5.6. موقع ويب أو منشور رقمي

هو كل المحتوى والتطبيقات المحددة على اسم النطاق. يمكن أن يكون نفس اسم النطاق:

- موقع الشركة،
- المتجر،
- بوابة الدعم،
- صفحة المستثمر،
- توثيق المطوّر،

- إصدارات مختلفة حسب البلد واللغة

قد تحمل. قد تختلف مالكو المحتوى ونطاقات الرقابة لهذه.

5.7. المنتج

شيء تجاري أو تقني يحمل وظيفة محددة أو نموذجًا أو إصدارًا. منتج:

- السعر،
- السلامة،
- الميزة التقنية،
- شهادة،
- ملاءمة المستخدم

لا يمكن استخدامها كخصائص عامة تابعة للشركة. قد تختلف الإصدارات ضمن نفس عائلة المنتج أيضًا.

5.8. الخدمة

نشاط يتم تقديمه ضمن نطاق محدد، مستخدم، بلد، طريقة التوصيل، ومسؤولية. الخدمة:

- قابلة للتسويق بواسطة العلامة التجارية،
- قابلة للفوترة بواسطة جهة قانونية أخرى

قابل للتسليم من قبل شريك مستقل. يجب أن يميز التدقيق من:

- باعه،
- سلمه،
- ضمنه،
- تحمل المسؤولية القانونية عنه

يجب أن يكون مفصولاً.

5.9. الشخص

هو مؤسس، أو مدير تنفيذي، أو موظف، أو مستشار، أو خبير، أو ممثل معترف به علنًا. الشخص:

- تعليمه،
- خبرته،
- رخصة،
- رأيه

لا يتم نقله تلقائيًا إلى المؤسسة. وبالمثل، لا يُعتبر جائزة المؤسسة أو شهادتها كفاءة مهنية للشخص.

5.10. الفرع أو الموقع الفعلي

هو نقطة خدمة فعلية تتبع المؤسسة الرئيسية أو تعمل بشكل منفصل. الفرع:

- جزء من الشركة الأم،
- كيان قانوني مستقل،
- امتياز،
- مشغل مستقل مرخّص

قد يكون. علاقة الموقع لا تظهر العلاقة القانونية بمفردها.

5.11. الامتياز

قد يكون مشروعًا مستقلًا يعمل تحت حق استخدام العلامة التجارية أو رخصة نموذج أعمال. يجب على مانح الامتياز:

- السمعة،
- السيطرة،
- التدريب

لا يضمن تلقائياً كل تصرف من قبل مشغل الفرنشايز. لا يمكن تعميم خطأ محلي لمشغل الفرنشايز تلقائياً على العلامة التجارية العالمية بأكملها. يجب الإبلاغ عن العلاقة مع نطاقها.

5.12. الموزع، التاجر أو البائع

الطرف الذي يبيع أو يوزع المنتج أو الخدمة في سوق محددة، سواء كان مستقلاً أو تابعاً. لا يتم نقل ترخيص الشركة المصنعة أو سيطرتها أو امتثالها المؤسسي تلقائياً إلى الموزع. كما أن أداء خدمة العملاء للبائع ليس هو نفس مقياس جودة منتج الشركة المصنعة.

5.13. المنصة وبائع المنصة

السوق أو المنصة والبائعون المستقلون على المنصة كيانان منفصلان. المنصة:

- الدفع،
- القائمة،
- البنية التحتية التقنية

يمكن توفيرها. البائع:

- المنتج,
- التسليم،
- الضمان،
- علاقات العملاء

يمكن أن يكون مسؤولاً. إذا جمع الذكاء الاصطناعي هذه الأدوار، فإن تخصيص المسؤولية بشكل غير صحيح قد يحدث.

5.14. برنامج، شارة أو شهادة

مقدمة من مؤسسة:

- برنامج تدريبي،
- علامة المطابقة،
- شهادة،
- عضوية،
- شارة

يمكنه حمل سجلات كيانات منفصلة. المؤسسة التي تنشر البرنامج ليست هي نفسها المؤسسة المشاركة في البرنامج. ميزة الشارة لا تترجم إلى جميع ميزات المؤسسة التي تحمل الشارة.

6. رسم بياني للكيان

قد لا يكون سجل كيان واحد كافيًا لشرح العلاقات المعقدة. لهذا السبب، يجب إنشاء رسم بياني للكيان لكل تدقيق.
وجهات النظر:

G_E = (V, R)

هنا:

- V: عقد الكيان
- R: العلاقات الاتجاهية بين العقد

يجب تعريف العلاقات بأفعال صريحة.

6.1. الأنواع الأساسية للعلاقات

إلى الحد ذي الصلة، يمكن استخدام العلاقات التالية:

- يمتلك — يمتلك
- ownedBy— مملوك من قبل
- يسيطر — يسيطر
- controlledBy— يتم التحكم فيه من قبل
- يشغل — يشغل
- operatedBy— تُدار من قبل
- ينشر — ينشر
- publishedBy— تُنشر بواسطة
- يُصنّع — يُصنّع
- manufacturedBy— تُصنع بواسطة
- providesService— تقدم خدمة
- serviceDeliveredBy— يتم تقديم الخدمة بواسطة
- licencesBrandTo— ترخص العلامة التجارية لـ
- franchisesTo— يمتلك حقوق الامتياز لـ
- distributesFor— يوزع لـ
- sells — يبيع
- employs — يوظف
- foundedBy— تم تأسيسه بواسطة
- memberOf— عضو في
- certifiedBy— مُعتمد في النطاق المحدد بواسطة
- auditedBy— تم تدقيقه بواسطة
- partneredWith— لديه شراكة في النطاق المحدد مع
- customerOf— عميل لـ
- supplierTo— مورد لـ
- formerlyKnownAs— كان يُعرف سابقًا باسم
- predecessorOf— سلف لـ

- successorTo— هو خلف لـ
- acquiredBy— تم الحصول عليه بواسطة
- mergedInto— تم دمجه في
- ceasedOn— توقف عن العمل في

ينبغي عدم استخدام الأفعال الغامضة مثل 'مرتبط بـ' أو 'شريك' أو 'يعمل معًا' بمفردها قدر الإمكان.

6.2. اتجاه العلاقة

العلاقات لها اتجاه. 'A يقدم التكنولوجيا لـ B.' ليست نفسها مثل: 'B يقدم التكنولوجيا لـ A.' قد ينتمي علامة تجارية إلى شركة. الشركة لا تنتمي إلى العلامة التجارية. قد يتعاون العميل مع وكالة. الوكالة لا تنتمي إلى العميل. إذا لم يتم تحديد الاتجاه، قد يقوم النظام بنقل المسؤولية والسلطة بشكل غير صحيح.

6.3. مدة العلاقة

يجب تسجيل كل علاقة مالية بالقدر الذي تكون فيه ذات صلة:

- تاريخ البدء،
- تاريخ الانتهاء،
- حالة النشاط،
- آخر تاريخ تحقق

يجب الإدراج. الشراكة السابقة ليست شراكة نشطة. الموظف السابق ليس تنفيذياً حالياً. الموزع الذي تم استخدامه من قبل قد لا يكون ممثلاً معتمدًا اليوم.

6.4. نطاق العلاقة

قد تقتصر الشراكة على:

- منتج محدد،
- البلد،
- حملة،
- تاريخ،
- حدث

لا يمكن تعميم أن تعاون مشروع واحد ينطبق على جميع الأنشطة التجارية.

7. يُحظر افتراض نقل الخصائص

وجود علاقة بين الكيانات لا ينتج عنه تلقائيًا نقل صفة من واحدة إلى أخرى. أحد المبادئ الأساسية لهذا القسم هو:

> يُحظر نقل الخاصية بشكل افتراضي؛ ومع ذلك، يمكن السماح به بعلاقة واضحة ونطاق وتوقيت وأدلة محددة.

7.1. نقل الترخيص

ترخيص الشركة الأم:

- إلى شركة فرعية،
- إلى امتياز،
- إلى موزع,
- إلى علامة تجارية

لا يتم النقل تلقائيًا. يجب فحص النطاق الرسمي للرخصة.

7.2. نقل الشهادة

لا يمكن نقل شهادة المنتج إلى كامل عائلة المنتجات. لا يمكن تعميم شهادة الجودة لمكتب واحد على المجموعة العالمية. قرار الشركة بالمطابقة لا يثبت الخبرة الشخصية لموظفيها.

7.3. نقل عدد الموظفين

لا يمكن استخدام إجمالي المجموعة كعدد موظفي شركة محلية. لا يمكن اعتبار شبكة الوكالات أو قاعدة المستقلين عدد الموظفين المباشرين.

7.4. نقل عدد العملاء

لا يمكن كتابة عدد عملاء المجموعة الرئيسية كعدد عملاء علامة تجارية جديدة. قد لا يتطابق عدد المستخدمين على منصة مع عدد العملاء النشطين لمنتج معين.

7.5. تحويل تاريخ التأسيس

سنة بدء المؤسس في الصناعة ليست تاريخ تأسيس الشركة. لا يمكن للشركة استخدام تاريخ تأسيس علامة تجارية قديمة تستحوذ عليها كتاريخ تأسيس قانوني لها. يجب عرض تاريخ العلامة وتاريخ الكيان القانوني بشكل منفصل.

7.6. النجاح وإحالة الحالة

المشروع الذي يقوم به موظف لدى صاحب عمل سابق ليس دراسة حالة للشركة الجديدة. نجاح شركة تم الاستحواذ عليها، دون توضيح التكامل والنطاق، لا يُعتبر النجاح المباشر للمجموعة الجديدة.

7.7. نقل الأهلية

تم منح الأهلية لصفحات لغة معينة من اسم المجال:

- لجميع المجموعة الشركة،
- للمنتجات الجديدة،
- لأسماء مجالات أخرى،
- لصفحات مستقبلية

لا يتم نقلها.

8. أربع طبقات ملكية منفصلة

في التمثيل الرقمي، يمكن أن تحمل كلمة "المالك" معانٍ متعددة. يجب التمييز بين أربع طبقات.

8.1. الملكية القانونية

الكيان هو المالك بموجب القانون أو التسجيل الرسمي.

8.2. ملكية العلامة التجارية

هو المالك لحقوق العلامة التجارية أو الهوية التجارية.

8.3. التحكم في الواجهة الرقمية

هو الطرف الذي يسيطر فعليًا على اسم النطاق أو الموقع الإلكتروني أو الحساب الاجتماعي أو التطبيق. يمكن أن يكون المالك القانوني والمتحكم الفني مختلفين.

8.4. ملكية المحتوى والادعاء

إنها الجهة المسؤولة عن دقة الادعاء بالمادة المنشورة. قد تدير الوكالة الموقع. ومع ذلك، فإن المالك الحقيقي للادعاء بالسعر أو الترخيص قد يكون المؤسسة العميلة. المزود الفني لا يصبح مالك المادة للادعاء لمجرد أنه نشر المحتوى.

9. من يبيع الخدمة، ومن يقدمها؟

هذا أحد أهم الفروق بين الكيانات في تقييم GEO. في سلسلة الخدمة، قد تختلف الأطراف التالية:

- الشخص الذي يستخدم العلامة التجارية
- الشخص الذي يعلن
- الشخص الذي يحضر العرض
- الشخص الذي يوقع العقد
- الشخص الذي يصدر الفاتورة
- الشخص الذي يقدم الخدمة بالفعل
- الشخص الذي يقدم الضمان
- الشخص الذي يعالج البيانات
- الشخص الذي يتعامل مع الشكوى
- الشخص الذي يتحمل المسؤولية القانونية

قد يقوم الذكاء الاصطناعي بتعيين جميع هذه الأدوار لطرف واحد بمجرد ذكر اسم العلامة التجارية. هذا الخطأ شائع بشكل خاص في:

- السياحة،
- التمويل،
- الصحة،
- الاستشارات،
- امتياز،
- المنصات،
- الوكالات،
- التعليم

يمكن أن تؤدي الحقول إلى عواقب مالية. يجب أن يجيب كل سجل خدمة على هذه الأسئلة: مع من يبرم المستخدم العقد؟ من يتلقى الدفع؟ من يقوم بالخدمة؟ من يتحمل المسؤولية المهنية أو القانونية؟ من المسؤول عن الالتزامات المتعلقة بالضمان والإرجاع؟ ما دور العلامة التجارية؟ هل هناك مزودون من طرف ثالث؟ هذا التمييز يحدد أيضًا الكيان الذي ينطبق عليه درجة GEO.

10. حدود الكيان

حدود الكيان هي الخط الذي يحدد الميزات والعلاقات والواجهات التي يشملها الكائن الذي تم تدقيقه والتي لا يشملها. يمكن تعريف حدود الكيان في أربع طبقات.

10.1. جوهر الهوية

إنها الهوية الأساسية للكيان التي يجب أن تبقى دون تغيير:

- الاسم القانوني الموحد
- نوع الكيان
- الهوية القانونية أو المؤسسية
- النشاط الرئيسي
- الملكية الأساسية
- الحالة النشطة

10.2. بيئة مُسيطر عليها

يتم التحكم بها مباشرة من قبل الكيان:

- أسماء النطاقات،
- التطبيقات،
- العلامات التجارية،
- المنتجات،
- المستندات،
- الملفات الرسمية

هذا المجال. التحكم لا يعني أن جميع الميزات تندمج في هوية واحدة.

10.3. بيئة متصلة

الشركات التابعة هي شركات محلية، امتيازات، وشركاء متحكم بهم. تحتفظ بهوياتها المنفصلة.

10.4. بيئة العلاقة الخارجية

العملاء، الموردون، وسائل الإعلام، العضويات، مزودو التكنولوجيا، والشركاء المستقلون. الكيانات في هذه البيئة ليست جزءاً من الكيان التي تخضع للتدقيق. إنها موضوعات العلاقة.

11. حصر حدود الكيان دون رؤية النتائج

لا يمكن توسيع أو تضييق موضوع التدقيق بناءً على النتائج. السلوكيات التالية محظورة:

- إضافة نتائج المجموعة الإيجابية إلى العلامة التجارية
- استبعاد شركة تابعة فاشلة من النطاق
- نشر نتيجة منتج جيد عبر جميع عائلة المنتجات
- إعلان نسخة اللغة السيئة ككيان منفصل
- ترك الخطأ الحرج للإدارة المحلية مع جعل العلامة التجارية العالمية تظهر نظيفة
- جعل الأدلة المحلية القوية هي دليل الشركة العالمية

نطاق الكيان:

- قبل بدء التدقيق،
- قبل توزيع صياغات المُدخل،
- قبل رؤية النتائج

يجب أن يتم إصدار النسخة أولاً.

12. سجل الكيان القانوني

يجب إنشاء سجل كيان قانوني لكل كيان تم تدقيقه. يوضح هذا السجل الكائن الذي يقوم الحكام والأنظمة بتقييمه.

12.1. الحد الأدنى من الحقول

الهوية

NOMOS معرف الكيان الاسم الرسمي نوع الكيان الاسم القانوني الاختصارات الأسماء المحلية الرسمية الأسماء السابقة

الوضع القانوني والشركي

الولاية القضائية رقم التسجيل تاريخ التأسيس الحالة النشطة الشركة الأم الشركات التابعة

العلامات التجارية والمنتجات

العلامات التجارية المملوكة العلامات التجارية المرخصة المنتجات الخدمات النسخة أو الفروق الطرازية

الواجهات الرقمية

الأسماء النطاقية الرسمية الأسماء النطاقية المحلية التطبيقات الملفات الرسمية السجلات القابلة للقراءة آليًا

المسؤولية

المطالب صاحب المحتوى مالك الواجهة التقني مالك تقديم الخدمة المسؤول القانوني

الوقت

إصدار السجل بدء النفاذ نهاية الصلاحية تاريخ آخر تحقق التغييرات التاريخية

النطاق

الدول التي تم تدقيقها اللغات التي تم تدقيقها المنتجات التي تم تدقيقها الكيانات والواجهات خارج النطاق

12.2. حالات الكيانات

يجب أن يحتوي كل سجل كيان على واحدة من الحالات التالية:

نشط

غير نشط

تاريخي

تم استبداله

دمج

تم الاستحواذ عليه

تم حله

مؤقت

متنازع عليه

مجهول

لا يمكن حل الهوية المجهولة أو المتنازع عليها بشكل نهائي بصمت.

13. مستويات إثبات الهوية

يمكن تصنيف مستوى التأكيد على هوية سجل كيان وفقًا للمصادر المستخدمة.

ID-0— هوية غير معروفة

يوجد اسم أو اسم نطاق فقط. لم يتم التحقق من المطابقة القانونية أو التنظيمية.

ID-1— تصريح من الطرف الأول

تعرّف المؤسسة نفسها بطريقة معينة. إنها ذات قيمة. إنها ليست تحققًا مستقلاً.

ID-2— هوية تنظيمية مدعومة

سجلات الطرف الأول واسم المجال والسجلات الموثوقة الأخرى متوافقة مع بعضها البعض.

ID-3— هوية مؤكدة بسجل رسمي

تم التحقق من الكيان القانوني باستخدام السجلات الرسمية أو الموثوقة المناسبة.

ID-4— هوية تم التحقق منها عبر عدة واجهات وعلاقات

الهوية والعلامة التجارية والمنتج واسم المجال والعلاقات المؤسسية متوافقة عبر سجلات موثوقة متعددة.

ID-5— الهوية المحفوظة على مر الزمن

تم التحقق من الهوية والعلاقات في فترات مختلفة، وتم الحفاظ على سجل التغيير. هذه المستويات ليست درجة الأهلية النهائية. إنها تشير إلى ثقة الأدلة في سجل الكيان.

14. كيانات مختلفة بنفس الاسم

قد توجد كيانات مختلفة بنفس الاسم أو بأسماء مشابهة. هذا الوضع يخلق تصادمًا بين الكيانات. أمثلة على التمييز:

- شركات بنفس الاسم
- مؤسسات تستخدم نفس الاختصار
- نفس اسم العلامة التجارية في دول مختلفة
- شركة قديمة وجديدة
- اسم العلامة التجارية مع الاسم الشخصي
- اسم الشركة مع اسم المنتج
- العلامة التجارية مع كلمة عامة

لا يمكن للمراجعة الاعتماد فقط على مطابقة الاسم.

14.1. حقول حل الاصطدام

الولاية القضائية نوع الشركة القانونية اسم النطاق العنوان الفعلي تاريخ التأسيس المؤسس أو المدير المنتج والخدمة مالك العلامة التجارية رقم التسجيل الاسم باللغة المحلية

14.2. ادعاء غامض

إذا أشار الادعاء إلى عدة كيانات معقولة، فهناك ثلاثة خيارات: يتم توضيح الادعاء. يُصنَّف بشكل منفصل كاختبار لحل الكيانات. من المتوقع أن يطالب النظام بالتوضيح. لا يمكن إضافة الإجابة المستخلصة من الادعاء الغامض بصمت إلى الدرجة الرئيسية للدقة.

15. كيان، أكثر من اسم نطاق واحد

شركة:

- نطاق عالمي،
- نطاقات الدولة المحلية،
- نطاقات المنتج،
- نطاق المستثمر،
- نطاق الدعم،
- النطاقات القديمة

يمكن استخدامها. ليست جميع هذه النطاقات تقع تلقائيًا تحت نفس نطاق التدقيق. لكل نطاق:

- المتحكم،
- مالك المحتوى،
- المستخدم المستهدف،
- اللغة،
- نطاق المنتج،
- حالة النشاط

يجب تحديده. السجل الصحيح في مجال واحد لا يصحح تلقائياً الخطأ في مجال آخر.

16. اسم نطاق، أكثر من كيان واحد

على اسم نطاق واحد:
- الشركة الأم،
- العلامات التجارية المختلفة،
- الشركات المحلية،
- المنتجات،
- الشركاء،
- البائعون

يمكن أن يحدث ذلك. في هذه الحالة، يجب تحديد ملكية الكيان بناءً على الرابط الإلكتروني أو عائلة المحتوى. مثال:

الواجهة	دور الوجود
الصفحة الرئيسية	العلامة التجارية العالمية
علاقات المستثمرين	الشركة الأم القانونية
صفحة الأسعار المحلية	منظمة المبيعات المحلية
سوق	البائعون المستقلون
صفحة الوظائف	الكيانات صاحب العمل المحددة
بوابة الدعم	مشغل المنتج أو الخدمة

لا يعني اسم نطاق واحد مسؤولية واحدة.

17. هوية الكيان المؤقتة

قد تتغير هوية الكيان مع مرور الوقت. أمثلة على الحالات:
- تغيير اسم الشركة
- انتقال العلامة التجارية إلى شركة أخرى
- الاندماج
- الاستحواذ
- القسم
- إنهاء النشاط
- تغيير اسم النطاق
- نقل المنتج إلى شركة أخرى
- رحيل المؤسس
- تغيير المشغل المحلي

لهذا السبب، تسجيل الكيان يعتمد على الوقت. يمكن اعتباره كـ E(t).

17.1. الهوية التاريخية والهوية الحالية

العلاقة التي كانت صحيحة في الماضي لا يمكن نقلها إلى الحاضر. مثال: "تم تشغيل X بواسطة Y في 2022." هو تاريخي. الجملة التالية مختلفة: "يتم تشغيل X بواسطة Y."

17.2. نقل الميزة بعد الاستحواذ

الاستحواذ:

- الملكية،
- بعض الكيانات،
- قد تنقل بعض العقود

. ومع ذلك:

- جميع الإنجازات التاريخية،
- جميع الشهادات,
- جميع علاقات العملاء،
- جميع التراخيص

لا يتم نقلها تلقائيًا إلى الكيان الجديد. يجب توضيح الميزات التي يتم نقلها قانونيًا وعمليًا.

17.3. تاريخ أسماء النطاقات

قد يحمل اسم النطاق آثارًا للمحتوى أو إشارات من المالك السابق. لا يُعتبر المالك الجديد تلقائيًا قد حصل على السلطة السابقة. إذا دمج نظام الذكاء الاصطناعي الكيانات القديمة والجديدة، يحدث التباس في الكيانات التاريخية.

18. هوية الكيان متعددة اللغات

يجب ربط أسماء الكيان في لغات مختلفة بنفس الهوية. ومع ذلك، ليست كل الترجمات أسماء رسمية. يجب التمييز بين الأنواع التالية:

- الاسم القانوني
- اسم العلامة التجارية
- الاسم الرسمي المحلي
- النقل الصوتي
- الترجمة الوصفية
- الاختصار
- الاسم السابق
- الاسم المستخدم بشكل شائع

مثال: لا يجب أن يكون الوصف التركي للعلامة التجارية هو الاسم القانوني أو الرسمي التركي للعلامة التجارية. يجب أن يشير السجل الآلي بوضوح إلى هذه الأنواع.

19. يمكن أن تتغير الدولة والاختصاص القضائي حدود الكيان

تحت علامة تجارية عالمية، قد توجد كيانات قانونية أو تشغيلية مختلفة في بلدان مختلفة. عندما يسأل المستخدم، "هل تعمل الشركة X في ألمانيا؟" يجب أن يكون الجواب مميزاً:

- وجود العلامة التجارية العالمية،
- الكيان القانوني المحلي
- علاقة الموزع
- القدرة على تقديم الخدمات عن بعد

«العلامة التجارية معروفة في البلاد» ليست هي نفسها: «شركة قانونية محلية تقدم خدمات.»

19.1. أدوار الكيانات حسب الدولة

لكل دولة، بالقدر الذي ينطبق:

- البائع القانوني
- مُصدر الفاتورة
- متحكم البيانات
- مزود الخدمة
- الدعم
- الموزع
- الامتياز
- حامل الترخيص
- ممثل العلامة التجارية

يجب تحديده.

19.2. الصفحة باللغة العامة ليست دليلاً على وجود شركة محلية

وجود موقع إلكتروني باللغة الألمانية لا يثبت وجود كيان قانوني أو مكتب فعلي في ألمانيا. الصفحة باللغة الإنجليزية لا تخلق سلطة قانونية عالمية. اللغة ليست سلطة قضائية.

20. مؤشر الوجود الاصطناعي APPLE.COM

عرض المنهجية التركيبية / تم إعداد المثال التالي فقط لعرض طريقة حل الكيانات. إنه ليس التدقيق الفعلي لشركة آبل، ولا مراجعة الهيكل المؤسسي الحالي، ولا النتيجة الحقيقية لأي منتج ذكاء اصطناعي. المُدخل: "ما نوع الشركة Apple.com؟"

20.1. مرساة المُدخل

apple.com

20.2. الأهداف المرشحة

اسم النطاق والموقع الإلكتروني علامة آبل التجارية الكيان المؤسسي القانوني وظيفة المتجر الإلكتروني عائلة المنتجات إذا لم يتم تحديد الهدف قبل بدء التدقيق، قد يقوم المقيمون بتقييم كائنات مختلفة.

20.3. تثبيت الكيان الهدف المقترح

يمكن تعريف الهدف من الاختبارات التركيبية على النحو التالي: "الكيان التكنولوجي الرئيسي للشركة المرتبط باسم النطاق القانوني apple.com، والذي يمثل هوية الشركة وأنشطتها الأساسية." هذا التعريف:

- لا يشمل تلقائيًا المواصفات التقنية للمنتجات الفردية،
- الشركة المحلية لكل دولة،
- أسعار المتاجر،
- النصائح الاستثمارية

تلقائيًا.

20.4. الاستجابة التركيبية أ

“Apple.com هو موقع إلكتروني للشركة العالمية مرتبط بالمنتجات التكنولوجية والخدمات الرقمية.” التقييم: لا يتطابق تمامًا مع المؤسسة وفقًا لاسم النطاق. قد يترك هوية الشركة غير مكتملة. النشاط الرئيسي قد يكون صحيحًا بشكل عام. قد يكون هناك نقص في الهوية المادية.

20.5. استجابة تركيبية ب

“Apple.com هو الموقع الرسمي للشركة المسماة iPhone.” التقييم: تم دمج المنتج والشركة معًا. تم حل الكيان المستهدف بشكل خاطئ. هناك خطأ في الهوية.

20.6. استجابة تركيبية ج

“Apple.com هو مجرد متجر مستقل يبيع الهواتف عبر الإنترنت.” التقييم: وظيفة اسم النطاق ضيقة بشكل مفرط. الكيان المؤسسي ونطاق الأنشطة مشوه. قد تكون صفة المتجر المستقل غير صحيحة.

20.7. الاستجابة التركيبية د

“Apple.com هو أحد الواجهات الرئيسية للمعلومات المؤسسية ومنتجات علامة Apple التجارية؛ يجب أيضًا تحديد الكيان القانوني الذي يدير الموقع وجانب المبيعات المحلي وفقًا للبلد.” التقييم: يتم فصل اسم النطاق والعلامة التجارية والدور القانوني المحلي. تحليل الكيان الهدف أكثر انضباطًا. يجب إثبات الادعاءات المادية الأخرى للاستجابة بشكل منفصل. يوضح هذا المثال:

> يمكن أن تتركز الاستجابة على الكيان غير الصحيح حتى لو كانت تحتوي على العديد من الكلمات الصحيحة.

دقة الهوية تأتي قبل دقة المحتوى.

21. حالة كيان مركب اصطناعي

حالة اصطناعية — ليست مؤسسة حقيقية

لننظر إلى علامة تجارية عالمية باسم أسترون. الهيكل:

- أسترون هولدينغز المحدودة — الشركة الأم
- أسترون ترافيل — العلامة التجارية
- أسترون تركيا تورزم المحدودة — الشركة المحلية
- ميرا لخدمات الوجهات — شريك عمليات محلي مستقل
- نادي أسترون أنطاليا — عملية امتياز
- asteron.example — اسم نطاق عالمي
- tr.asteron.example — الواجهة التسويقية التركي

يسأل المستخدم: "ما هي الخدمات التي تقدمها أستيرون في تركيا؟" رد الذكاء الاصطناعي: "تدير شركة Asteron Holdings Ltd. جميع الجولات في أنطاليا نفسها، ولديها 400 موظفًا في تركيا، وAsteron Club Antalya مسؤول مباشرة عن رضا العملاء في أنطاليا." حزمة الواقع: الشركة الأم تمتلك العلامة التجارية. الشركة المحلية هي جانب المبيعات والتعاقدات. تقوم Mira بتنفيذ جزء من العمليات. Asteron Club Antalya هي امتياز مستقل. موظفو 400 يمثلون المجموع الكلي للمجموعة. الامتياز مسؤول مباشرة عن رضا العملاء على مستوى الأعمال. الأخطاء في الرد: تم تقديم الشركة الأم على أنها المسؤولة عن تقديم الخدمة. يتم نقل العدد الإجمالي لموظفي المجموعة كما لو كان هو السعة الخاصة بالشركة المحلية. يتم تعيين مسؤولية الامتياز مباشرةً للشركة الأم. يتم دمج العلامة التجارية والكيان القانوني والشريك التشغيلي. قد يكون الرد قد سمى العلامة التجارية بشكل صحيح. ومع ذلك، فإن مخطط الكيان غير صحيح.

لذلك، لا يمكن أن يعتمد مؤشر GEO فقط على ذكر الاسم أو الطلاقة العامة للجملة.

22. نتائج حل الكيان

بالنسبة لكل ملاحظة، يمكن تسجيل نتيجة حل الكيان بأحد الحالات التالية.

ER-0— لم يتم التقييم

لم يتم تقييم حل الكيان.

ER-1— الهدف الدقيق

تم حل الكيان المستهدف بشكل صحيح وواضح.

ER-2— تطابق مقبول على مستوى العلامة التجارية

لم يتم تقديم اسم قانوني، لكن العلامة التجارية والكيان المؤسسي المستهدف تم تحديدهما بشكل صحيح بشكل جوهري.

ER-3— جزئي أو غير محدد بشكل كاف

كان الرد يقترب من المنطقة الصحيحة لكنه يفتقر إلى الهوية القانونية أو الجغرافية أو التشغيلية.

ER-4— التداخل بين الكيانات المتعددة

تم دمج عدة كيانات منفصلة ككائن واحد.

ER-5— الكيان غير الصحيح

تم استهداف شخص آخر أو شركة أو منتج أو اسم نطاق.

ER-6— غامض

لا يمكن تحديد الكيان الذي تم حله من الرد.

ER-7— توضيح مناسب

فعلاً، المُدخل غامض، والنظام طلب التوضيح بشكل صحيح. لا يجب اعتبار طلب التوضيح فشلاً.

23. نجاح حل الكيان يختلف عن دقة التمثيل

يمكن للنظام تحليل الكيان الصحيح ولكن قد يقدم معلومات غير صحيحة عنه. قد يذكر نظام آخر بعض الحقائق المادية بشكل صحيح ولكنه يربطها بـ الكيان غير الصحيح. هناك أربع حالات أساسية:

تحليل الكيان	المحتوى المادي	النتيجة
صحيح	صحيح	مرشح تمثيل صحيح قوي
صحيح	غير صحيح	كائن صحيح، سمة غير صحيحة
غير صحيح	صحيح جزئيًا	معلومات صحيحة مرتبطة بـ الكيان غير الصحيح
غير صحيح	غير صحيح	تحليل تمثيل كامل

لهذا السبب، قد يكون تحليل الهوية بوابة منفصلة. على وجه الخصوص، الإجابة التي تحتوي على ارتباك شديد مع كيان آخر لا يمكن إنقاذها بمعدل دقة واقعية مرتفع.

24. ربط قرار المطابقة بالكيان

يجب أن يتم تأسيس قرار المطابقة بالتنسيق التالي:

```
[معرّف الكيان] + [نوع الكيان] + [النطاق] + [الواجهات الرقمية] + [الدولة واللغات] + [الوقت] + [إصدار المعيار]
```

نموذج المثال: "بالنسبة لـ [العلامة التجارية/الكيان القانوني] المحدد بمعرّف الكيان NOMOS NGE-XXXX؛ تم تقييم الصفحات العامة باللغتين التركية والإنجليزية على النطاق المحدد، ضمن نطاق المنتجات والخدمات المحدد، في [التاريخ] وفقًا لـ [إصدار المعيار]." ثم يجب التأكيد بوضوح: "لا ينطبق هذا القرار تلقائيًا على الشركات الفرعية، الفروع، المنتجات خارج النطاق، النطاقات الأخرى، واللغات التي قد تُضاف في المستقبل."

25. الأحكام الإلزامية المعيارية

CH03-N01

يجب على كل تدقيق جغرافي تحديد كيان مستهدف فريد قبل عرض النتائج.

CH03-N02

لا يمكن افتراض أن المرساة المُدخلة هي نفسها الكيان المستهدف.

CH03-N03

يجب تمييز أسماء النطاقات والعلامات التجارية والكيانات القانونية والمنتجات والخدمات والهويات الشخصية عن بعضها البعض.

CH03-N04

لا يمكن نقل خاصية من كيان إلى كيان ذي صلة آخر دون دليل واضح ونطاق محدد.

CH03-N05

لا يمكن تطبيق ترخيص الشركة الأم أو شهادتها أو عدد موظفيها أو عملائها أو نتيجة الامتثال تلقائيًا على الشركات التابعة.

CH03-N06

يجب فصل الادعاء على مستوى العلامة التجارية عن الطرف الذي يقدم الخدمة فعليًا وهو المسؤول قانونيًا.

CH03-N07

يجب تسجيل صاحب الامتياز ومشغل الامتياز والموزع والتاجر وبائع المنصة ككيانات منفصلة.

CH03-N08

يجب أن تحمل كل علاقة كيان ملموس معلومات عن الاتجاه والنطاق والوقت.

CH03-N09

لا يمكن استخدام العلاقات مثل "شريك" و"عميل" و"خبير" و"شركة مجموعة" وما شابه ذلك بشكل غامض.

CH03-N10

يجب استخدام مناطق حل التصادم للكيانات المختلفة التي تحمل نفس الاسم.

CH03-N11

إذا كان من المقرر تضمين ادعاء غير مؤكد في النتيجة الرئيسية، يجب تحديد طريقة التعامل مع عدم اليقين مسبقًا؛ وإلا يجب توضيح الادعاء.

CH03-N12

لا يمكن توسيع نطاق الكيان أو تضييقه بعد رؤية النتائج من أجل زيادة النتيجة الإيجابية.

CH03-N13

لا يمكن استخدام الكيانات والعلاقات التاريخية كما لو كانت تحمل حالة الكيان الحالية.

CH03-N14

الشراء أو الدمج أو نقل العلامة التجارية لا ينقل تلقائيًا الخصائص السابقة إلى الكيان الجديد.

CH03-N15

يمكن تطبيق قرار التطابق فقط على وحدة التطابق المحددة في القرار.

CH03-N16

لا يمكن للمنتج الجديد أن يرث تلقائيًا التوفر الحالي للغة أو اسم النطاق أو الشركة التابعة أو الدولة.

CH03-N17

يجب تقييم نجاح حل الكيان والدقة الواقعية بشكل منفصل.

CH03-N18

لا يمكن جعل دمج الهوية المادية مع الكيان غير الصحيح غير مرئي ضمن المتوسط الواقعي العام.

CH03-N19

يجب أن تكون سجلات الكيان الرسمية قابلة للإصدار وأن تحتفظ بالتغييرات مع مرور الوقت.

CH03-N20

يجب أن يكون هناك مالك إنساني أو مؤسسي مسؤول لسجل الكيان.

26. أشكال الفشل

CH03-F01 — اعتبر اسم النطاق كشركة

يتم تقديم العنوان الرقمي ككيان قانوني أو تشغيلي.

CH03-F02 — دمج العلامة التجارية مع الشركة القانونية

يتم استخدام اسم العلامة التجارية بدلًا من الكيان القانوني الذي يتحمل العقود والمسؤوليات.

CH03-F03 — اعتبر المنتج كشركة

يتم حل اسم المنتج كشركة أو هوية علامة تجارية.

CH03-F04 — نقل خصائص المؤسس إلى الشركة

يتم جعل عمر المؤسس أو خبرته أو جائزته أو المشروع السابق سمة من سمات الشركة.

CH03-F05 — نقل خصائص المجموعة إلى الشركة المحلية

يتم تخصيص بيانات العملاء أو الموظفين أو الإيرادات المجمعة للأعمال المحلية.

CH03-F06 — نقل الترخيص من خلال العلاقة

يتم استخدام ترخيص الشركة الأم أو المنتج أو الموظف أو الشريك باسم كيان آخر.

CH03-F07 — توحيد الامتياز والعلامة التجارية الرئيسية

يتم تقديم أداء ومسؤولية الأعمال المحلية مباشرة كعلامة تجارية عالمية.

CH03-F08— الجمع بين المنتج والبائع

يُعتبر مصنع المنتج والبائع المحلي أو الموزع ككيان واحد.

CH03-F09— دمج المنصة والبائع

تُحمّل منصة السوق مسؤولية منتج البائع أو الضمان أو التسليم.

CH03-F10— اعتبار العلاقة كنقل ميزة

يُسوّق وضع العميل أو العضوية أو استخدام التكنولوجيا أو الشراكة بشكل مضلل على أنه نقل للسلطة.

CH03-F11— تحديث الهوية التاريخية

يُستخدم الاسم القديم أو الملكية أو الشراكة أو المدير كما لو كانت الواقع الحالي.

CH03-F12— المطالبة بالماضي بعد الاستحواذ

جميع النجاحات السابقة للكيان المكتسب تُنسب إلى المالك الجديد بدون أي تفسير.

CH03-F13— دمج الكيانات التي تحمل نفس الاسم

يتم اعتبار شركات أو أفراد مختلفين ككيان واحد بسبب تشابه الأسماء.

CH03-F14— تقسيم الكيان بشكل مصطنع

يتم إعلان لغة أو منتج أو عملية محلية فاشلة ككيان منفصل للحفاظ على تقييم إيجابي.

CH03-F15— توسيع الكيان بشكل مصطنع

يتم توزيع نتائج العلامة التجارية أو المنتج الناجح عبر المجموعة بأكملها.

CH03-F16— النظر في نطاق الاستحقاق

يُستخدم نتيجة التدقيق الضيقة للشركة بأكملها أو المجموعة أو مجموعة المنتجات.

CH03-F17— احتساب اختصاص اللغة

يُستخدم وجود نسخة اللغة كدليل على القدرة القانونية للشركة أو الخدمة في ذلك البلد.

CH03-F18— احتساب التحكم الفني كملكية

يتم عرض الوكالة أو المطور كمالك قانوني أو تجاري للمؤسسة لأنهم يديرون اسم النطاق.

CH03-F19— ترك المدعي غير محدد

لا يمكن تحديد من المسؤول عن معلومات السعر أو الترخيص أو القدرة غير الصحيحة.

CH03-F20— تفسير مرساة الطلب بعد النتيجة

يتم اختيار الكيان المستهدف الذي سيجعل الاستجابة تظهر ناجحة لاحقًا.

27. إجراء التدقيق

الخطوة 1— تسجيل مرساة الطلب

يتم تسجيل أي من التالي لدى المستخدم بدقة:

- الاسم،
- اسم المجال،
- علامة تجارية،

- المنتج,
- الشخص

التعبير المستخدم بالكامل.

الخطوة —2 تحديد الكيان المستهدف

يتم ملاحظة الكيان الذي يقيسه التدقيق فعليًا. يتم تصنيف الهدف كالتالي:

- كيان قانوني،
- علامة تجارية،
- المنتج,
- الخدمة،
- حزمة الكيان

يجب تسجيل الفئة المطبقة بشكل صريح.

الخطوة —3 إنشاء معرف الكيان القانوني

الاسم القانوني نوع الكيان الهوية القانونية مجال الاختصاص الأسماء المستعارة حالة الوقت مسجلة.

الخطوة —4 بناء رسم كيان الكيان

يتم استخراج العلاقات الموجهة بين الشركة الأم والشركات التابعة والعلامة التجارية والمنتج واسم المجال والأفراد والشركاء.

الخطوة —5 تحديد أدوار المسؤولية

يتم فصل صاحب الادعاء وصاحب الواجهة الرقمي وصاحب تقديم الخدمة والمسؤول القانوني وصاحب الأدلة.

الخطوة —6 مراجعة خطر نقل السمة

ردًا على ذلك:

- رخصة،
- شهادة،
- الموظف،
- العميل،
- تاريخ التأسيس،
- نجاح،
- الامتثال

هل تم نقلها إلى الكيان غير الصحيح؟

الخطوة —7 التحقق من الهوية الزمنية

الإجابة:

- الاسم السابق،
- المالك السابق،
- الشراكة السابقة المنتهية،
- الموظف السابق،

- الحالة قبل الاستحواذ

هل يتم استخدامه؟

الخطوة 8— التحقق من الأدوار الجغرافية والمحلية

هل الأدوار العالمية للعلامة التجارية، الشركة المحلية، الموزع، والامتياز صحيحة؟

الخطوة 9— تصنيف الكيان الذي حُدد من قبل النظام

تم تعيين الحالة المناسبة بين ER-1 و ER-7.

الخطوة 10— وحدة توافق التثبيت

لتطبيق القرار النهائي:

- الكيان،
- المجال،
- المنتج,
- اللغة،
- البلد،
- الوقت

يتم كتابة الحدود.

الخطوة 11— تحديد الادعاء العامة

تضيق النتيجة إذا تم تعميمها لمؤسسة أو مجموعة أوسع من الكيان المحدد.

الخطوة 12— تسجيل طريقة التغيير والاعتراض

إذا تم الإبلاغ عن خطأ في الهوية أو العلاقة:

- من سيتحقق،
- من سيغير،
- أي نسخة سيتم إنشاؤها

يتم تحديدها.

28. الدليل الضروري

هوية الكيان القانونية؛ الاسم القانوني ونوع الكيان؛ السجل الرسمي أو التحقق المكافئ؛ ملكية العلامة التجارية؛ السيطرة على اسم النطاق؛ العلاقات بين الشركة الأم والشركة التابعة؛ ملكية المنتجات والخدمات؛ سجلات الأعمال المحلية والموزعين؛ علاقات الامتياز؛ طرف التعاقد والفوترة؛ المسؤولية عن تقديم الخدمة؛ حامل الترخيص والشهادة؛ مؤسس وأدوار تنفيذية؛ الأسماء النشطة والتاريخية؛ سجلات الاستحواذ والاندماج والنقل؛ الدولة والاختصاص القضائي؛ السجلات الرسمية للغة والترجمة الصوتية؛ رسم بياني للكيان؛ المطالب؛ مالك المحتوى؛ مالك الواجهة الفني؛ نطاق التدقيق؛ الكيانات والواجهات خارج النطاق؛ نسخة سجل الكيان؛ تاريخ آخر تحقق؛ وشخص أو منظمة مسؤولة.

29. قائمة تدقيق التدقيق

هل تم تسجيل مرساة المُدخل بوضوح؟ هل تم تحديد الكيان المستهدف قبل ملاحظة النتائج؟ هل تم فصل الكيان القانوني واسم النطاق؟ هل تم فصل العلامة التجارية وصاحب العلامة التجارية؟ هل تم فصل المنتج والشركة؟ هل تم فصل سمات المؤسس عن سمات المؤسسة؟ هل العلاقات بين الشركة الأم والشركة الفرعية صحيحة؟ هل الامتيازات والموزعون مميزون؟ هل تم تحديد الأطراف التي تبيع وتقدم الخدمة؟ هل تم فصل المدعي والناشر التقني؟ هل التراخيص والشهادات تتبع الكيان الصحيح؟ هل تم ذكر أرقام الموظفين والعملاء والإيرادات ضمن النطاق الصحيح؟ هل تم فصل الأسماء والعلاقات التاريخية عن الحالة الحالية؟ هل تم التحقق من وجود كيانات مختلفة بنفس الاسم؟ هل تم ربط مجالات متعددة وواجهات المستخدم بالكيان الصحيح؟

هل تم ربط الأسماء متعددة اللغات بنفس الهوية الرسمية؟ هل الأدوار الخاصة بالدولة والولاية القضائية واضحة؟ هل تم تصنيف الكيان بشكل منفصل نتيجة لحل الكيان؟ هل وحدة الامتثال محددة بوضوح؟ هل تمتد إلى الكيانات خارج نطاق القرار؟ هل تم إصدار نسخ مؤرخة لسجل الكيان؟ هل هناك مسار للاعتراض وتصحيح التغييرات المتعلقة بالهوية؟

30. الاعتراضات والإجابات

الاعتراض 1— "المستخدم يعرف اسم العلامة التجارية، وليس الشركة القانونية."

صحيح. من الطبيعي أن يستخدم المستخدم العلامة التجارية في صياغة المُدخل. التدقيق لا يُجبر المستخدم على استخدام المصطلحات القانونية. ومع ذلك، التقييم في الخلفية:

- العلامة التجارية,
- المالك،
- مزود الخدمة،
- طرف التعاقد

يجب التمييز بينهم. يمكن إعطاء إجابة بسيطة للمستخدم. سجل التدقيق لا يجب أن يكون بسيطًا.

الاعتراض 2— "في شركة صغيرة، العلامة التجارية والشركة هما بالفعل نفس الشيء."

في بعض الهياكل الصغيرة:

- علامة تجارية واحدة،
- اسم نطاق واحد،
- مالك أعمال وحيد،
- مزود خدمة واحد

يمكن العثور عليه. في هذه الحالة، رسم كيان البيانات بسيط. ومع ذلك، يجب إجراء التسجيل صراحة. البساطة لا تعني الغموض.

الاعتراض 3— 'تتصرف مجموعة الشركات كعلامة تجارية واحدة.'

من منظور التسويق، قد يكون هناك علامة تجارية واحدة. من حيث المسؤولية القانونية والتشغيلية والأدلة، قد تكون هناك كيانات منفصلة. إذا كان من المفترض اتخاذ القرارات على مستوى المجموعة، يجب إظهار الميزات التي تم توحيدها بوضوح.

الاعتراض 4— "الكثير من التفاصيل يعقد درجة **GEO** بلا داعٍ."

الدرجة الصحيحة المعطاة لـ الكيان غير الصحيح غير مفيدة. حل الكيانات ليس عبئًا إضافيًا للتدقيق، بل هو شرط للصلاحية.

اعتراض 5— "الأنظمة الذكية بالفعل لا تولي اهتمامًا لهذه الاختلافات القانونية."

بالضبط لهذا السبب، يجب قياسه. نظام الذكاء الاصطناعي:

- علامة تجارية،
- شركة
- المنتج,
- الأعمال المحلية

فقدان الفرق بينهم هو خطأ التمثيل نفسه.

اعتراض 6— "إذا كانت الشركة الأم تتحكم في المجموعة بأكملها، يمكن نقل الخصائص."

علاقات السيطرة مهمة في بعض المجالات. ومع ذلك:

- رخصة،
- العقود,
- العميل،
- الموظف،
- مسؤولية البيانات
- الامتثال

تتطلب هذه الخصائص دليلًا منفصلًا للنقل. السيطرة ليست توحيد جميع الخصائص.

اعتراض 7— "هل إنجازات الشركة القديمة بعد الاستحواذ ليست جزءًا من المجموعة الجديدة؟"

يمكن تفسير ذلك في سياق تاريخي. ومع ذلك:

- خلال أي فترة تحقق النجاح،
- بواسطة أي فريق وكيان،
- قبل أو بعد الاستحواذ

يجب ذكر متى تم إنتاجه. الملكية الجديدة لا تغير التاريخ.

اعتراض 8— "الإرشاد واضح بالفعل؛ لقد كتب المستخدم اسم النطاق."

قد يكون اسم النطاق واضحًا. ومع ذلك: يجب أيضًا الإجابة على السؤال "ماذا يمثل اسم النطاق؟" يمكن أن يكون النطاق شركةً، متجرًا، بوابة منتج، أو منصة متعددة الكيانات.

اعتراض 9— "وصف الموزع المحلي كعلامة تجارية أسهل للمستخدم."

البساطة لا تتجاوز المسؤولية. يمكن تبسيط الرد كما يلي: "تقدم العلامة التجارية خدماتها في تركيا من خلال موزع محلي معتمد." هذا الرد مفهوم وصحيح في الوقت نفسه.

اعتراض 10— "كل هذه العلاقات تتغير مع الوقت."

هذا صحيح. لهذا السبب، سجل الكيان:

- بتاريخ،
- مؤرق,

- يجب أن يكون مدفوعًا بالتغيير

لا يقلل من الحاجة لتسجيل التقلبات. بل يزيدها.

القاعدة الشائعة للفصل 33

قبل توليد درجة GEO: نسأل، "كم عدد الإجابات الصحيحة؟" لكن السؤال الأكثر جوهرية الذي يجب طرحه قبل ذلك هو:

"إجابات من تحديدًا تتعلق هذه؟"

إذا اعتبرنا اسم النطاق كشركة، فإننا نقيس الكائن الخطأ. إذا اعتبرنا العلامة التجارية كشركة قانونية، فإننا نحدد المسؤولية بشكل غير صحيح. إذا اعتبرنا المنتج كشركة، فإننا نشوه ميزاته. إذا حولنا تجربة المؤسس إلى تاريخ الشركة، فإننا نبالغ في الماضي. إذا عدنا عدد موظفي المجموعة كقدرة محلية، فإننا نفخم توقعات الخدمة. إذا قامت الشركة الأم بنقل ترخيصها إلى امتياز، فإننا نوجه المستخدم إلى السلطة الخاطئة. إذا جعلنا الإنجازات السابقة بعد الاستحواذ ميزة للمالك الجديد، فإننا نعيد كتابة الحقيقة التاريخية. إذا قمنا بتمديد تدقيق اسم نطاق ضيق ليشمل الشركة بأكملها، فإننا نمحو نطاق الامتثال. لذلك، فإن تعريف الكيان ليس مجرد تنظيف للبيانات. إنه بداية للمسؤولية التمثيلية. يمكن للنظام أن يصف الكيان غير الصحيح بالمعلومات الصحيحة. هذا لا يزال تحريفًا. يمكن للنظام أن يصف الكيان الصحيح بالمعلومات الخاطئة. هذا أيضًا تحريف.

كلاهما مطلوب للوصول إلى GEO الصحيح:

> الكيان الصحيح. / السمة الصحيحة. / العلاقة الصحيحة. / النطاق الصحيح. / الوقت الصحيح.

لذلك، فإن قانون القياس الثالث لـ NOMOS هو:

> إذا لم يكن الكيان المقاس واضحًا، فإن دقة القياس أيضًا لن تكون واضحة.

القانون الرابع هو كما يلي:

> العلاقة ليست هوية.

القانون الخامس هو كما يلي:

> الملكية ليست نقل جميع الصفات.

قانونه السادس هو:

> الامتثال لا ينتشر إلى الكيانات المجاورة.

الطلب 3 القسم NOMOS

> لا تعطيني اسمًا وحده. / أرني الكيان الذي ينتمي إليه هذا الاسم.
>
> لا تعامل النطاق كشركة، أو العلامة التجارية كطرف متعاقد، أو المنتج كصانع.
>
> لا تستخدم ماضي المؤسس كعمر المؤسسة.
>
> لا تقم بنقل رخصة الشركة الأم إلى المنظمة التابعة أو الفرنشايز أو الموزع.
>
> لا تجعل عميل المجموعة عميل العلامة التجارية المحلية، أو تعتبر جائزة العلامة التجارية جائزة لجميع المنتجات.
>
> لا تصف التعاون بالملكية، أو العضوية بالاعتماد، أو استخدام التكنولوجيا بالشراكة.

لا تدمج الاسم القديم مع الكيان الجديد، أو الماضي المكتسب مع أداء اليوم.

لا توسع الكيان لتضخيم النتائج الإيجابية / لا تقسّم الكيان لإخفاء النتائج السلبية.

أولاً، قم بتحديد الهدف. / ثم سجّل اتجاه كل علاقة. / ثم أرفق كل صفة بمالكها الحقيقي. / ثم أضف الوقت والبلد ذي الصلة. / عندها فقط قرّر بشأن هذا الكيان وحده.

جملة ختام الفصل

السؤال الأول في تدقيق *GEO* ليس ما يقوله الذكاء الاصطناعي، بل تحديدًا من أو ما الذي تتعلق به كل عبارة.

النواة المعيارية

يجب على كل تدقيق **GEO**، قبل رصد النتائج، أن يحدد العناصر الآتية ويضبط إصداراتها: - مرتكز المُدخل، - الكيان المستهدف، - نوع الكيان، - سجل الهوية المرجعي، - مخطط العلاقات، - الحدود الزمنية والاختصاصية، - موضوع التدقيق، - ووحدة المطابقة. ويجب ألا تُعامل النطاقات والعلامات التجارية والكيانات القانونية ومجموعات الشركات والشركات التابعة والمنتجات والخدمات والأشخاص والامتيازات والموزعون والمنصات والبائعون بوصفها كيانات قابلة للتبادل. كما يجب ألا تنتقل السمات أو وثائق الاعتماد أو التراخيص أو العملاء أو الموظفون أو الأداء أو الصلاحيات أو حالة المطابقة بين كيانات مترابطة، ما لم تكن العلاقة والنطاق والزمن والأدلة تؤيد هذا الانتقال صراحة. ويجب تقييم الكيان الذي يحدده منتج الذكاء الاصطناعي بصورة مستقلة عن تقييم الدقة الواقعية للاستجابة. ولا يسري قرار المطابقة إلا على الكيان المحدد صراحة وعلى نطاق التدقيق المعلن، ولا يجوز أن يمتد تلقائيًا إلى الجهات التابعة أو المنتجات أو الاختصاصات القضائية أو اللغات أو النطاقات أو الإصدارات المستقبلية أو الكيانات المرتبطة.

طبقتان من الواقع

التمثيل الرسمي والواقع الموثَّق للكيان

الأسس المنهجية / K10; K11

حد الفصل

أقرت المادة 1 الحكم التالي:

> الاستجابة الواحدة ليست درجة *GEO*.

أضاف الفصل 2 ما يلي:

> التمثيل هو توزيع يحدث بين المستخدمين ومنتجات الذكاء الاصطناعي والدول واللغات والجلسات والأوقات.

ثم حدد الفصل 3 موضوع التوزيع:

> لا يمكن قياس الكيان حتى يتم حله.

والآن نأتي إلى السؤال الإلزامي التالي:

> وفقًا لأي سجل واقعي سنقيم ما إذا كان رد الذكاء الاصطناعي صحيحًا؟

أي مصدر يجب أن يحكم التدقيق: موقع الكيان الذي يتم تدقيقه، سجل قانوني، صفحة تسعير الشركة، تقارير مستقلة، هيئة الترخيص، مراجعات العملاء، بيانات أداء الشركة نفسها أم ادعاء المنافس؟ عندما تتناقض المصادر، أيها يجب قبوله كصحيح؟ إذا نشرت الشركة ادعاءً كاذبًا أو غير مدعوم وكرره الذكاء الاصطناعي بدقة، هل يكون النظام قد نجح؟ إذا كان الموقع الرسمي قديمًا لكن قدم الذكاء الاصطناعي إجابة مختلفة بناءً على أدلة خارجية حديثة وموثوقة، هل يكون النظام قد فشل؟ حتى يتم الإجابة على هذه الأسئلة:

- تمثيل صحيح،
- تمثيل غير صحيح،
- وفاء بالمصدر،
- الدقة الواقعية
- واقع الكيان

تُستخدم هذه المصطلحات أحيانًا كما لو كانت تشير إلى نفس المفهوم. تحل هذا الفصل الالتباس الناتج عن طريق التمييز بين طبقتين من الواقع:

الطبقة الأولى

طبقة التمثيل الرسمي

تُظهر ما يقوله الكيان عن نفسه على الواجهات القانونية التي يتحكم بها.

الطبقة الثانية

طبقة واقع الكيان الموثوق

تُظهر الأدلة الحالية المواد الادعاءات التي تدعم ضمن أي نطاق وزمن. هذان الطبقان مرتبطان، لكنهما ليسا متطابقين. الطبقة الأولى تجيب عن السؤال: «كيف يعرف الكيان نفسه؟» الطبقة الثانية تجيب عن السؤال: «أي جزء من هذا التعريف مدعوم بأي دليل؟» يحدد الكتاب الأول المعيار الصحيح، ويفصل الكتاب الثاني أشكال الانتهاك، ويُصمم نظام التدقيق لقياس وإثبات هذه الفروقات. تعمل هذه القسم على هذه المهمة التدقيقية على مستوى الادعاءات.

التحدي NOMOS

يحتوي موقع شركتكم على الإنترنت على الجملة التالية: "نحن الشركة الرائدة عالميًا في مجال GEO." يقرأ منتج ذكاء اصطناعي هذه الصفحة ويرد على المستخدم: "هذه الشركة هي الشركة الرائدة عالميًا في مجال GEO." هل يعكس الرد موقعكم بشكل صحيح؟ نعم. هل يعكس الرد الواقع المثبت بشكل صحيح؟

لا نعرف ذلك بعد. ربما:

- هناك أبحاث مقارنة،
- تم تحديد نطاق السوق،
- تم نشر معيار للريادة،

تم التحقق من النتيجة بشكل مستقل. ربما لا يوجد أي من هذه الأشياء. إذا كانت الذكاء الاصطناعي قد نقلت ادعاء الشركة نفسه: "تعرف الشركة نفسها بأنها الشركة الرائدة عالميًا في مجال GEO."، لأدت إلى الحفاظ على مصدر الادعاء. ومع ذلك، عندما قالت: "هي الشركة الرائدة عالميًا في مجال GEO.", قد تكون الشركة قد رفعت البيان إلى حقيقة مستقلة.

فكر في حالة أخرى. يذكر الموقع الرسمي للشركة: 'نحن نقدم الخدمات فقط في تركيا.' ومع ذلك، تُظهر السجلات المؤكدة أن الشركة بدأت العمل في ألمانيا والمملكة المتحدة منذ ستة أشهر، ولم يتم تحديث الموقع. بالاعتماد على السجلات القانونية والتشغيلية الحالية، ترد الذكاء الاصطناعي: 'تقدم الشركة الخدمات في تركيا وألمانيا والمملكة المتحدة.'

الذكاء الاصطناعي ليس مخلصًا تمامًا للموقع الرسمي. هل يمكن أن يكون مع ذلك أقرب إلى الواقع الحالي الذي تم التحقق منه؟ نعم. فكر في حالة ثالثة. يذكر موقع الشركة: 'نحن نعمل في 20 دول.' تسجل السجلات المستقلة والداخلية ثمانية دول فقط. يرد الذكاء الاصطناعي: 'تعمل الشركة في 20 دول.'

هذا الرد:

- مخلص للتمثيل الرسمي،
- مخالف للواقع المُتحقق منه

من الممكن. المواد الأولى من هذا القسم كما يلي:

> الإخلاص للسجل الرسمي ليس هو نفسه الدقة الواقعية.

ينص حكمه الثاني على:

> ما تقوله الشركة وما يمكن إثباته يجب تسجيله بشكل منفصل.

تنص الفقرة الثالثة منه على:

> تكرار الذكاء الاصطناعي لخطأ رسمي بشكل صحيح لا يُشكل تمثيلًا دقيقًا.

تنص المادة الرابعة على:

> انحراف الذكاء الاصطناعي عن سجل رسمي قديم ليس خطأ تلقائيًا إذا كان يحمل دليلًا موثوقًا.

1. غرض الفصل

الغرض من هذا القسم هو إنشاء طبقتين مرجعيتين منفصلتين لاستخدامهما في تقييم GEO-1000. يوضح القسم التمييزات التالية بشكل معياري:

- إعلان الوجود مقابل الحقيقة المثبتة
- المصدر الرسمي مقابل المصدر المستقل
- المصدر من الطرف الأول مقابل التحقق المستقل
- الولاء للمصدر مقابل الدقة المادية
- تكرار الادعاء مقابل التحقق من الادعاء
- السجل الرسمي الحالي مقابل السجل الرسمي القديم
- الصمت الرسمي مقابل المعلومات التي تم التحقق منها خارجيًا
- وجهة نظر الشركة مقابل الحكم على الواقع الخارجي
- الدليل الرسمي مقابل الادعاء الرسمي
- استقلالية المصدر مقابل صلة المصدر بالموضوع
- الخطأ الرسمي مقابل خطأ الذكاء الاصطناعي
- تضارب المرجع مقابل فشل النموذج
- عدم اليقين مقابل الفشل
- مصلحة الادعاءات مقابل صحة الادعاء
- وجود الأدلة وما إذا كانت الأدلة تدعم الادعاء بالفعل

بحلول نهاية هذا القسم، يجب أن يكون بالإمكان طرح سؤالين منفصلين لكل استجابة من الذكاء الاصطناعي:

> هل تنقل الاستجابة بشكل صحيح التمثيل الرسمي والموثوق للكيان؟

و:

> هل تعكس الاستجابة الواقع المادي الذي تحمله الأدلة القابلة للتحقق حالياً؟

قد تُفضي هذه الأسئلة إلى نفس الإجابة. وقد تُفضي أيضًا إلى إجابات مختلفة. وعندما تختلف، فإن دور التدقيق ليس إخفاء الاختلاف، بل شرحه.

2. الأحكام المعيارية المركزية

> يجب تقييم كل استجابة مادية للذكاء الاصطناعي مقارنة بالمرجعين المنفصلين دون تحويل: سجل التمثيل الرسمي وسجل واقع الكيان الذي تم التحقق منه.

استجابة:

- مخلص للتمثيل الرسمي،
- مناسب للواقع الذي تم التحقق منه,
- مناسبة لكليهما،
- بعيدة عن كليهما

يمكن أن يكون كذلك. لا يمكن لـ NOMOS دمج هذين النتيجتين بهدوء في قرار واحد صحيح–خاطئ. الأداء العالي في الطبقة الأولى لا يعوض عن الفشل في الطبقة الثانية. على وجه التحديد:

> يمكن لاستجابة الذكاء الاصطناعي التي تكرر بشكل صحيح ادعاءً لم يتم التحقق منه أو كاذب نُشر بواسطة الكيان نفسه أن تُظهر دقة التمثيل الرسمي؛ ومع ذلك، لا يمكنها إظهار دقة الكيان المؤكدة.

3. الطبقة الأولى

طبقة التمثيل الرسمية

طبقة التمثيل الرسمية هي السجل المزود بالإصدارات للبيانات المادية الحالية التي تنشرها عن نفسها على الوسائط المعتمدة التي تتحكم فيها الكيان أو التي تم تفويضها صراحة من قبله. تجيب هذه الطبقة على السؤال:

> كيف يريد الكيان أن يُفهم، وما المعلومات التي ينشرها رسمياً؟

تشمل هذه الطبقة:

- بيان الهوية القانونية،
- علاقة العلامة التجارية بالشركة،
- تعريفات المنتجات والخدمات،
- الأسعار المنشورة،
- دول الخدمة،
- شروط الدعم،
- السياسات الرسمية
- الحدود المكشوفة علناً،
- قد يحتوي على معلومات العملة الرسمية والإصدار

طبقة التمثيل الرسمية مهمة. لأن نظام الذكاء الاصطناعي يجب أن يفهم بشكل صحيح كيف كيان:

- defines itself, (يعرّف نفسه)،
- ما الخدمة التي يقدمها،
- ما السعر الذي ينشره،
- ما الشروط التي يقبلها

ومع ذلك، فإن طبقة التمثيل الرسمية ليست دليلاً مستقلاً تلقائيًا لكل ادعاءات. إن بيانًا على موقع المؤسسة نفسه، خاصة فيما يتعلق بالادعاءات الموجهة للعالم الخارجي مثل التفوق والأداء، يعرض البيان نفسه؛ لكنه لا يثبت صحته بمفرده.

4. ما هو المصدر الرسمي؟

لا يُعتبر المصدر رسميًا لمجرد أنه يحمل شعار المؤسسة. يجب أن يحتوي المصدر الرسمي على الأقل على الميزات ذات الصلة مما يلي:

- خاضع لسيطرة الكيان
- يخضع لعملية نشر معتمدة
- معتمد من قِبل صاحب المعلومات الواقعية
- رابط URL أصلي أو معرف السجل

- الإصدار
- تاريخ النشر أو التحقق
- سجل التغييرات
- الشخص أو المؤسسة المسؤولة
- النطاق
- حالة اللغة والموقع
- حالة الأرشفة أو التحديث

الرأي المنشور على حساب وسائط التواصل الاجتماعي الشخصي لموظف لا يُعتبر تلقائيًا بيانًا رسميًا للشركة. ملف مسودة من وكالة لا يُعد نشرًا رسميًا. بريد إلكتروني من مندوب مبيعات، إذا لم يكن مفوضًا، لا يُمثل سياسة عامة للشركة. ملف PDF قديم على موقع الشركة قد لا يحل محل السجل الرسمي الجديد.

5. ظهور التمثيل الرسمي

وفقًا للتدقيق المعني، قد تشمل الواجهات الرسمية:

- الموقع الإلكتروني الرئيسي للشركة
- إخلاءات المسؤولية القانونية
- صفحات المنتجات والخدمات
- صفحات التسعير والعروض
- مركز المساعدة الرسمي
- وثائق علاقات المستثمرين
- السياسات العامة والعقود
- النشرات الصحفية الرسمية
- البيانات المنظمة
- السجلات القانونية JSON
- مخرجات API الرسمية
- المحتوى الرسمي داخل التطبيق
- المواقع المحلية المصرح بها
- حسابات وسائل التواصل الاجتماعي الرسمية
- البيانات المقدمة للمؤسسات العامة
- سجلات الإصدار والسحب

قد تتناقض هذه السطوح مع بعضها البعض. في هذه الحالة، "التمثيل الرسمي" ليس واقعًا واحدًا وواضحًا. إنه مجموعة من السجلات التي يجب مراجعتها داخليًا.

6. سجل التمثيل الرسمي

يجب إنشاء سجل تمثيل رسمي لكل كيان تمت مراجعته. هذا السجل ليس نسخة من كل المحتوى الرسمي. إنه ملخص يمكن تتبعه للمزاعم والحدود الواقعية. يجب أن يتضمن كل سجل الحقول التالية إلى الحد الذي ينطبق:

- معرّف الادعاء
- النص الكامل للادعاء

- موضوع الادعاء
- نوع الادعاء
- المصدر الرسمي
- مالك المصدر
- اللغة والمنطقة
- تاريخ النشر
- تاريخ آخر تحقق
- فترة الصلاحية
- الحالة الرسمية
- النطاق
- القيود
- الإصدار السابق
- سبب التغيير
- المدعي
- الموافقة البشرية

بدون سجل تمثيل رسمي للكيان: لا يمكن اختبار الجملة "الذكاء الاصطناعي يعكس الموقع بدقة." بوضوح.

7. سلطة المصدر الرسمي تعتمد على نوع المزاعم

ليست كل المصادر الرسمية ذات وزن متساوٍ لكل ادعاء. يجب تقييم سلطة المصدر وفقًا لنوع الادعاء.

7.1. القرارات التي تكون تحت سيطرة الكيان نفسه

مثال:

- الخدمة التي يقدمها للجمهور
- السعر المنشور
- ساعات الدعم
- سياسة الإرجاع الرسمية
- قناة الاتصال
- إصدار المنتج المتاح للعامة

في هذه المناطق، يُعد المصدر الرسمي أساسيًا وغالبًا هو السجل المفوض الرئيسي. ومع ذلك، إذا كانت الصفحة الرسمية تتناقض مع سير العملية الفعلية، فإن الصفحة وحدها لا تكفي. على سبيل المثال:

- صفحة السعر 1,000 يورو،
- عملية الدفع 1,500 يورو

إذا كان يظهر، فهناك تناقض في طبقة العرض الرسمية.

7.2. الهوية القانونية والسلطة

مثال:

- تسجيل الشركة
- الرخصة

- السلطة المهنية
- ملكية العلامة التجارية
- الوضع الضريبي أو القانوني

يمكن لموقع الشركة أن يكشف عن هذه المعلومات. ومع ذلك، يمكن لسجل رسمي أو مؤسسة مفوضة أن يقدم تحققًا أقوى.

7.3. الادعاءات حول الأداء والنجاح

مثال:

- "98% من مشاريعنا ناجحة."
- "نحن نزيد إيرادات عملائنا بمقدار متوسط .40%"
- "لدينا أعلى معدل رضا."

قد يكون موقع الشركة على الإنترنت هو صاحب الادعاء. ولكن لدعم الادعاء:

- بيانات,
- طريقة
- المقام
- فترة،
- نتائج غير ناجحة،
- مراجعة مستقلة إذا أمكن

مطلوب.

7.4. ادعاءات التفوق والقيادة

مثال:

- زَعِيم عَالَمِي
- الأفضل
- الرقم واحد
- «المعيار الحقيقي الوحيد»
- «الشركة الأكثر موثوقية في الصناعة»

تصريح الموقع نفسه ليس مقارنة مستقلة. مثل هذه الادعاءات:

- يتطلب كونًا للمقارنة،
- معايير
- طريقة
- تاريخ،
- النطاق،
- الاستقلال

يجب تأسيس كلا الشرطين.

7.5. علاقات الطرف الثالث

مثال:

- عميل

- شريك
- كفيل
- العضوية
- الاعتماد
- شراكة تكنولوجية
- الوكالة

يمكن أن يكون البيان الرسمي الأحادي الجانب ذا قيمة. ومع ذلك، قد يحتاج إلى التحقق من قبل الطرف الآخر في العلاقة أو من قبل سجل موثوق.

7.6. نتائج المستخدم

مثال:

- رضا العملاء
- النتيجة الفعلية للاستخدام
- معدل الإرجاع
- معدل الشكاوى
- الشراء المتكرر

بيانات الشركة في هذه المجالات تُعد دليلاً من الطرف الأول مهم. ومع ذلك:

- اختيار البيانات،
- عدم الاستجابة,
- المنهجية،
- تضارب المصالح

يجب الإفصاح عنها.

8. دقة التمثيل الرسمي

المفهوم الذي يشير إلى مدى توافق استجابة الذكاء الاصطناعي مع طبقة التمثيل الرسمي:

دقة التمثيل الرسمي

المصطلح الإنجليزي الرسمي هو:

دقة التمثيل الرسمي

هذا القياس يطرح الأسئلة التالية: هل يقوم الذكاء الاصطناعي بتفسير الكيان بشكل صحيح؟ هل ينقل المنتجات والخدمات الرسمية بدقة؟ هل يشوه السعر الرسمي؟ هل يزيل الحدود الرسمية؟ هل يوسع التغطية الرسمية للبلد؟ هل يعرض رأي الشركة كحقيقة مستقلة؟ هل يستخدم معلومات رسمية قديمة على أنها حالية؟ هل يخفي التناقضات بين السجلات الرسمية؟ هل يضيف ادعاءات مادية غير موجودة في المصدر الرسمي؟ هل يحافظ على جودة المصدر؟
إن دقة التمثيل الرسمي هي قياس لمدى صحة نقل الكيان للواقع الذي نشره. هذا القياس:

> إنه لا يُظهر مدى صحة تصريحات الكيان، بل مدى دقة الذكاء الاصطناعي في نقل ما يقوله الكيان.

9. الطبقة الثانية

طبقة واقع الكيان المؤكد

طبقة الواقع للكيان المُحقق هي سجل تقييم مُرقّم للمعلومات المادية التي كانت موجودة في تاريخ قياس محدد ويدعمها دليل مناسب لنوع الادعاء. هذه الطبقة تجيب على السؤال:

> ما هي التصريحات المتعلقة بالكيان المدعومة فعليًا بأية أدلة؟

تشمل هذه الطبقة:

- السجلات القادمة من الشركة نفسها،
- السجلات العامة الرسمية،
- المصادر المستقلة,
- سجلات العقود أو المعاملات،
- وثائق الترخيص،
- السجلات التاريخية المُحققة،
- الأدلة المضادة
- النتائج السلبية،
- عدم اليقين،
- النزاعات المتبقية

يمكن تقييمها معًا.

10. "الواقع المُحقق" ليس الواقع المطلق

هذا البروتوكول لا يزعم: "نحن نعرف بيقين جميع الحقائق المطلقة عن مؤسسة." الواقع المُحقق للكيان هو:

- معتمد على الأدلة،
- معتمد على الزمن،
- معتمد على نطاق.
- قابلة للتغيير مع الأدلة الجديدة،
- يمكن الطعن عليها،
- يمكن سحبها،

قد تحمل عدم اليقين. التعريف الصحيح هو:

> ضمن التاريخ والنطاق المحددين، سجل الكيان الأقوى والقابل للتحقق الذي تحمله الأدلة المتاحة.

لذلك، سجل الواقع المُحقق:

- مؤرق,
- مصدر،
- قابل للاعتراض،
- قابل للتعديل،
- واعٍ بعدم اليقين

يجب أن يكون. NOMOS لا يمكن أن يعتبر نفسه صاحب الحقيقة. فهو يصنف فقط ما الذي تسمح به الأدلة وإلى أي مدى.

11. سجل الواقع للكيان المُحقق

سيتم تفصيل البنية التقنية الكاملة لهذه الطبقة في فصل لاحق من الكتاب. على المستوى الأساسي، يجب أن تتواجد الحقول التالية لكل ادعاء واقعي:

- معرّف الادعاء
- الكيان الموضوع
- العلاقة أو السمة
- القيمة
- النطاق
- الدولة أو الاختصاص القضائي
- الوقت
- المصادر
- أنواع المصادر
- التحقق من المصدر
- الاستقلالية
- الأدلة المضادة
- كفاية الأدلة
- حالة التحقق
- عدم اليقين
- قرار الحكم
- حالة الاستئناف
- تاريخ آخر تحقق
- نسخة السجل

لا يمكن إنشاء هذا السجل بناءً على الأدلة الإيجابية فقط. يجب أيضًا الاحتفاظ بالسجلات التي تضيق الواقع أو تتناقض مع الادعاء.

12. حالات التحقق

يمكن تصنيف كل ادعاء مادي بإحدى الحالات التالية:

VR-0— غير مقيم

لم يتم تقييم الادعاء بعد.

VR-1— ادعاء رسمي فقط

الادعاء قائم فقط على بيان المؤسسة الرسمي الخاص بها. لم يُثبت أنه زائف. لا يوجد تحقق مستقل أو إضافي كافٍ.

—VR-2 دليل من الطرف الأول مدعوم

الادعاء مدعوم من قبل المؤسسة من:

- المعاملة،
- العملية،
- العقود,
- بيانات,
- السجل

مدعوم بالأدلة. المصدر من الطرف الأول.

—VR-3 مُؤكد

الادعاء مدعوم بعدة سجلات مناسبة ومتسقة. ليس على جميع المصادر أن تكون مستقلة.

—VR-4 تم التحقق منه بشكل مستقل

يتم دعم الادعاء بواسطة مصدر مستقل مناسب أو مراجعة حول الموضوع. الاستقلالية وحدها لا تعني الكفاية. كما أن صلة المصدر بالموضوع مطلوبة أيضًا.

—VR-5 مستمر متعدد المصادر

يتم دعم الادعاء بشكل قوي من قبل مصادر موثوقة من أنواع مختلفة ضمن وقت ونطاق محددين.

VR-C — متناقض

يتعارض الادعاء مع دليل أقوى أو أكثر ملاءمة.

VR-U — لم يُحل بعد

الأدلة غير كافية أو متناقضة. لا يمكن إصدار حكم موثوق.

VR-T — محدود الزمن

يتم دعم الادعاء فقط للفترة المحددة.

VR-S — محدود النطاق

الدعوى مدعومة فقط للمنتج المحدد أو الدولة أو المستخدم أو نطاق العملية. يمكن أن تحمل الدعوى أكثر من حالة في نفس الوقت. مثال:

VR-4+ VR-T + VR-S

يمكن أن يشير إلى أن الادعاء تم التحقق منه بشكل مستقل؛ ومع ذلك، فإنه يحمل قيوداً زمنية ونطاقية.

13. الاستقلال ليس هو نفس الملاءمة

المصدر المستقل ليس دائمًا أفضل مصدر. قد يقدر مدون سعر الشركة الحالي، في حين أن صفحة التسعير الرسمية للشركة قد تكون المصدر الأكثر مباشرة وملاءمة. وبالمثل، عندما تعلن الشركة على موقعها الإلكتروني أنها 'القيادة في الصناعة'، قد يكون البحث المستقل مع طريقة معلنة هو الدليل الأنسب. لذلك، تقييم المصدر يتطلب أكثر من مجرد معرفة ما إذا كان المصدر مستقلًا. هل هو موثوق لهذا النوع من الادعاءات؟ ما مدى قربه من الادعاء؟ هل نطاقه مناسب وحديث؟ من يتحكم فيه؟ هل تم الكشف عن طريقته؟ هل يتناول الأدلة المضادة؟ هل يمكن تدقيقه؟ هل هناك تضارب في المصالح؟ الاستقلال بمفرده لا يضمن الدقة.

كون المصدر من الطرف الأول لا يجعله عديم القيمة أيضًا. الشيء المهم هو:

امتلاكه السلطة المناسبة وتأهيل الأدلة للادعاء الذي يحمله المصدر.

14. خريطة ملاءمة الادعاء-المصدر

الجدول أدناه هو خريطة أساسية عامة. قد تختلف متطلبات الأدلة المحددة حسب القطاع ومستوى المخاطر.

نوع الادعاء	المصدر أو الدليل الأساسي	دور الموقع الرسمي	الحاجة إلى تحقق إضافي
الاسم القانوني للشركة	السجل الرسمي المعتمد	يعرض الإعلان	مطلوب بشكل عام
ملكية العلامة التجارية	تسجيل العلامة التجارية، العقد، سجل الشركة	يعرض الإعلان	مطلوب في حالة البيانات المالية
السعر المنشور الحالي	السجل المرجعي للسعر / الدفع	عادة رئيسي	قابل للتحقق من خلال سير العمل
نطاق الخدمة	سجلات العمليات، العقود، الصفحة الرسمية	البيان الرئيسي	أدلة إضافية على التسليم الفعلي
الرخصة	جهة الترخيص أو السجل	يعرض الإعلان	مطلوب
علاقة العميل	العقد، موافقة العميل، السجل الثنائي	يمكن أن يكون ادعاءً أحاديًا	مطلوب في حالة البيانات المالية
معدل النجاح	البيانات الخام، الطريقة، المقام، التدقيق	ينشر الادعاء	مطلوب بشدة
القيادة	الطريقة والبيانات المقارنة	إعلان ذاتي	يتطلب مقارنة مستقلة أو قوية
رضا العملاء	مسح تمثيلي، طريقة، عدم الاستجابة	يمكنه شرح النتيجة	المنهجية مطلوبة
تاريخ التأسيس	السجلات الرسمية والوثائق التاريخية	يعرض الإعلان	يجب فصل التاريخ القانوني وتاريخ العلامة التجارية
شراكة	بيان أو عقد ثنائي	إعلان أحادي	مطلوب بشكل عام
الشهادة	المؤسسة المصدرة للشهادة والنطاق	يعرض الإعلان	مطلوب
موظف نشط	سجل الموارد البشرية أو العقد، الملف العام	يمكنه عرض الإعلان	التحقق من الدور والتاريخ
نتيجة المستخدم	سجل المعاملة والنتيجة	يصف الحالة	السياق والطريقة والإذن المطلوب
"أفضل خيار"	المعايير المحددة للمستخدم والمقارنة	وجهة نظر التسويق	يتطلب مقارنة قوية

هذا الجدول ليس تصنيفًا عالميًا للمصادر. إنه يوضح مدى ملاءمة الادعاء والأدلة.

15. مصفوفة الادعاءات ذات الطبقة المزدوجة

يتم طرح سؤالين منفصلين لكل ادعاء مالي: هل تم نشرها رسميًا؟ ما هو وضعها في السجل المؤكد للحقائق؟ هذان السؤالان يشكلان الحالات الأساسية التالية.

15.1. رسمي ومؤكد

الادعاء: موجودة في سجل رسمي ومدعومة بالأدلة المناسبة. مثال: السعر الموجود في صفحة السعر الرسمية يتم تأكيده أيضًا في تدفق المعاملات. هذا هو أحد أقوى أنواع التمثيل الأساسي. إذا نقل الذكاء الاصطناعي الادعاء بالنطاق الصحيح والتوقيت المناسب، يحدث النجاح في كلا الطبقتين.

15.2. رسمي ولكن معلن من قبل الذات فقط

الادعاء رسمي. ومع ذلك، لا يوجد تحقق إضافي. مثال: "نحن الشركة الأكثر ابتكارًا في الصناعة." يمكن للذكاء الاصطناعي نقل الادعاء كالتالي: "تصف الشركة نفسها بأنها الشركة الأكثر ابتكارًا في الصناعة." لا يمكنه رفعها إلى: "هي الشركة الأكثر ابتكارًا في الصناعة."

15.3. رسمي ولكن متناقض

الادعاء موجودة في السجل الرسمي. تتعارض مع أدلة أكثر ملاءمة أو محدثة. مثال: الموقع يقول 20 دولة. تظهر السجلات التشغيلية ثماني دول. إذا قام الذكاء الاصطناعي بتكرار الموقع بالضبط، قد تكون الأمانة الرسمية عالية. الدقة المُحققة منخفضة. هذا الوضع:

نقل الخطأ الرسمي بأمانة

يمكن تصنيفه كالتالي.

15.4. غير رسمي ولكنه مُحقق

لم يتم العثور على الادعاء في الموقع الرسمي. ومع ذلك، تم التحقق منه من خلال مصادر موثوقة أخرى. المثال: لم يُضاف التغيير الأخير في الملكية القانونية بعد إلى الموقع الرسمي. لكنه يظهر في السجل المعتمد. يمكن للذكاء الاصطناعي نقل هذه المعلومات بدقة مع المصدر والتوقيت الصحيحين. قد تكون التمثيلات الرسمية منخفضة الدقة أو مختلفة. قد تكون الدقة الواقعية المثبتة عالية.

15.5. غير رسمي وغير مؤكد

الادعاء: غير موجود في السجل الرسمي، ويفتقر إلى الأدلة الخارجية الكافية. يجب ألا يقوم الذكاء الاصطناعي بنقل هذا وكأنه حقيقة.

15.6. المادة محذوفة رسميًا

تنشر الشركة الادعاء الإيجابي على صفحتها الرسمية. ومع ذلك، فهي لا تنشر القيد الضروري لفهمه بشكل صحيح. مثال: "نحن نقدم خدمات في أوروبا." ومع ذلك، تُقدَّم الخدمة في بلدين فقط وعن بُعد. إذا كررت الذكاء الاصطناعي الجملة العامة فقط، فإنه يمكن أن يظل مخلصًا للنص الرسمي. ومع ذلك، قد ينتج عنه نتائج مضللة جوهريًا. لذلك، يجب على طبقة التمثيل الرسمي تسجيل ليس فقط الجمل المنشورة ولكن أيضًا الإغفالات الجوهرية.

15.7. النزاع غير محلول

السجلات الرسمية والخارجية تتناقض مع بعضها البعض. ولا أي منها قوي بما يكفي. النتيجة الصحيحة:

غير محلول

ينبغي أن يكون. لا يمكن توقع أن يقوم الذكاء الاصطناعي بإصدار حكم قطعي.

16. جدول تقييم ذو طبقتين

الوفاء بالتمثيل الرسمي	دقة الواقع المؤكدة	تعليق أساسي
عالي	عالي	تمثيل مخلص وموثوق
عالي	منخفض	نقل الخطأ الرسمي بأمانة
منخفض	عالي	تمثيل تصحيحي يختلف عن السجلات الرسمية ولكنه مدعوم بالأدلة
منخفض	منخفض	فشل التمثيل العام
غير واضح	عالي	المصدر الرسمي متضارب؛ التحقق الخارجي قوي
عالي	غير واضح	تم نقل البيان الرسمي بشكل صحيح؛ التحقق من الواقع غير كافٍ
غير واضح	غير واضح	مرجع غير كافٍ للقرار

يقرّر هذا الجدول مبدأً مهماً:

> عدم امتثال الذكاء الاصطناعي للسجل الرسمي ليس دائمًا فشلًا.

وكذلك:

> امتثال الذكاء الاصطناعي للسجل الرسمي ليس دائمًا نجاحًا.

17. قاعدة للحفاظ على حالة المصدر

يجب أن يحافظ رد الذكاء الاصطناعي ليس فقط على المحتوى ولكن أيضًا على الحالة المعرفية للمحتوى. الجمل التالية ليست هي نفسها:

- "الشركة هي المنظمة الجغرافية الرائدة عالميًا."
- تصف الشركة نفسها بأنها المنظمة الجغرافية الرائدة في العالم.
- "تدعي الشركة أنها المنظمة الجغرافية الرائدة في العالم بناءً على أبحاثها المنشورة التي لم يتم التحقق منها بعد بشكل مستقل."
- في المقارنة المستقلة المحددة، احتلت الشركة المرتبة الأولى.

كل جملة تحمل حالة دليل مختلفة. الذكاء الاصطناعي:

- الإعلان الذاتي كنتيجة مستقلة
- بيان تسويقي كاكتشاف علمي
- بيان صحفي كخبر تحريري
- جائزة برعاية باعتبارها اعتمادًا مستقلاً
- بيانات الشركة كمدقق خارجي

هذا العلاج محظور بموجب القاعدة التالية:

قاعدة حماية الاقتباس وحالة المصدر

يجب أن تظل حالة المصدر صريحة.

18. الارتقاء المعرفي

تعزيز حالة مصدر الادعاء في استجابة الذكاء الاصطناعي يسمى:

الارتقاء المعرفي

أمثلة على التحولات:

النص المصدر	تعزيز غير صحيح بواسطة الذكاء الاصطناعي
"وفقًا لبيانات الشركة نفسها"	"مثبت"
"أفاد به عملاء محددون"	"يعمل لجميع العملاء"
"تمت ملاحظته في الدراسة التجريبية"	"فعال بشكل عام"
"تُعرف الشركة نفسها كقائدة"	"رائدة في الصناعة"
"لم يتم القيام بأي تحقق مستقل"	تم إزالة القيد تمامًا
"متوفر في بعض الأسواق"	"متوفر عالميًا"

هذه التحولات ليست ملخصات. إنها إنتاج لادعاءات جديدة وأقوى.

19. الصمت الرسمي

غياب المعلومات على الصفحة الرسمية للشركة قد لا يعني:

- المعلومات غير صحيحة,
- المعلومات ممنوعة،

لا يمكن للذكاء الاصطناعي استخدام هذه المعلومات. يمكن للصمت الرسمي أن يتخذ ثلاثة أشكال مختلفة.

19.1. الصمت غير المادي

قد تكون المعلومات خارج المُدخل وقرار المستخدم. سواء استخدمها الذكاء الاصطناعي أم لا، فهذا ليس ضروريًا.

19.2. معلومات خارجية تكميلية شرعية

قد تكون معلومات غير موجودة على الموقع الرسمي لكنها مؤكدة من مصدر موثوق. مثال:

- سجل قانوني
- حدث تاريخي
- قرار تنظيمي عام
- بحث مستقل

يمكن للذكاء الاصطناعي استخدام هذه المعلومات مع السياق المناسب وحالة المصدر.

19.3. صمت حد مادي

قد لا تكون الجهة قد أفصحت عن حد يؤثر على قرار المستخدم في تمثيلها الرسمي. مثال:

- الترخيص صالح فقط في بعض الدول
- السعر يظهر وكأنه لا يحمل رسومًا إضافية
- الخدمة متاحة فقط لأنواع معينة من العملاء

- بيانات النجاح تأتي من حالات مختارة فقط

في هذه الحالة، قد يحمل التمثيل الرسمي نفسه نتيجة مراجعة. تكرار الذكاء الاصطناعي للصمت قد يقلل من مصداقية الواقع.

20. نقص جوهري في التمثيل الرسمي

قد لا تكون الجملة الرسمية خاطئة بشكل مباشر. ومع ذلك، فإن إزالة الحدود اللازمة يمكن أن يجعلها مضللة. مثال: "نحن نعمل في أوروبا." قد يتم تقديم السياق المادي:

- فقط في دولتين،
- فقط عن بُعد،
- بدون مكتب محلي،
- فقط للعملاء من الشركات

يجب على طبقة التمثيل الرسمية تسجيل هذين المجالين معًا:

- الادعاء الرئيسي المنشور
- الحدود المادية اللازمة لفهم الادعاء بشكل صحيح

إذا كان رد الذكاء الاصطناعي يحمل الجملة الرئيسية لكنه يزيل الحدود:

- أمين من حيث الكلمات،
- مضلل من حيث الحكم

قد يكون.

21. الخطأ خطأ حتى لو كان في صالحك

عندما يصوّر منتج الذكاء الاصطناعي كيانًا ما على أنه أكثر قوة أو أكبر أو أكثر سلطة مما هو عليه فعليًا، يصبح الصمت من قِبَل المنظمة مسألة قياس وحوكمة. قد يُصوَّر شركة محلية على أنها عمل عالمي؛ أو شركة استشارية على أنها جهة تنظيمية؛ أو التقييم الذاتي على أنه شهادة مستقلة؛ أو عملية تمتد عبر ثمانية بلدان على أنها تعمل في 20. حتى عندما يكون الخطأ في صالح المنظمة، يجب تسجيله كخطأ في طبقة الواقع المُحقق. التحذير المؤسس للكتاب الثاني يضع نفس الحد: يجب تصحيح الخطأ التجاري الجذاب. عندما تستغل المنظمة عن عمد خطأً مفيدًا من الذكاء الاصطناعي:

- الالتزام بالتصحيح،
- نزاهة التمثيل،
- الوعد التجاري.
- الحوكمة

يمكن أيضًا فحصه من حيث

22. تعارض مرجعي

أحيانًا المشكلة ليست في استجابة الذكاء الاصطناعي، بل في سجلات المرجع. مثال: تعرض الصفحة الرئيسية سعرًا مختلفًا. تعرض وثيقة PDF سعرًا مختلفًا. تعرض صفحة الدفع سعرًا آخر. البيانات المنظمة تحتوي على سعر قديم. في هذه الحالة، قد لا يكون هناك تمثيل رسمي صحيح واحد. مثال آخر: يشير السجل القانوني إلى ملكية واحدة.

يذكر موقع الشركة ملكية أخرى. تواريخ التحديث غير مؤكدة. في هذه الحالة، قد يحمل سجل الواقع الموثق أيضًا تباينًا.

23. قاعدة العيب المرجعية

> إذا كانت سجلات المراجع متناقضة بشكل جوهري أو غير كافية، فلا يمكن تحويل هذا الغموض تلقائيًا إلى فشل في منتج الذكاء الاصطناعي.

الحالات الصحيحة:

تعارض_المراجع

دليل غير كافٍ

غير محلول

تقسيم الوقت

تقسيم النطاق

ممكن. يجب على التدقيق أن يفصل بين السؤالين: "هل قدّم الذكاء الاصطناعي إجابة غير صحيحة؟" و: "هل كان هناك مرجع ضروري للواقع لكي يقدم الذكاء الاصطناعي الإجابة الصحيحة؟" إذا تركت مؤسسة ما سجلاتها الرسمية متناقضة، يمكننا قياس أي سجل يختاره النموذج. ومع ذلك، لا يمكننا إخفاء خلل المرجع مع توقع إجابة مؤسسية واحدة ونهائية من النموذج.

24. تثبيت السجل الرسمي

يجب إصدار نسخ من كلا طبقتي المرجع لـ GEO-1000 قبل بدء موجة القياس:

سجل الممثل الرسمي

أي الصفحات؟ أي نسخ لغوية؟ أي تاريخ؟ أي ادعاءات؟ أي حالة أرشيفية؟

سجل واقع الكيان الموثّق

أي دليل؟ أي الأدلة المضادة؟ ما هي حالات التحقق؟ ما هي الاعتراضات؟ ما هو التاريخ؟ إذا كان هناك تغيير جوهري أثناء القياس: يمكن إيقاف الموجة. يمكن تقسيم السجل إلى نسخة جديدة. يمكن تحليل الملاحظات السابقة واللاحقة بشكل منفصل. يُحظر مزج النتائج تحت مرجع واحد.

25. عبء الإثبات مع قوة الادعاء

كلما ازداد الادعاء، يزداد أيضًا عبء الإثبات. السلم أدناه هو تمثيل بنّاء.

C0 — ادعاء هوية بسيط

"X هي شركة استشارية." مطلوب سجل الكيان والنشاط الصحيح.

C1 — ادعاء نطاق

"X تقدم خدمات في ألمانيا." مطلوب سجل تشغيل أو عقد أو تسليم.

C2 — ادعاء أداء كمي

"90% من المشاريع يتم إنجازها في الوقت المحدد." مطلوب المقام، الفترة، الطريقة، والنتائج غير الناجحة.

C3 — ادعاء مقارن

"إنه أسرع من المنافسين." يتطلب وجود كون المقارنة وقياس مكافئ.

C4 — ادعاء التفوق

"إنه الرائد في الصناعة." تتطلب الصناعة والمعايير والجغرافيا والزمن والطريقة.

C5 — الضمان أو التوفير الشامل

"يضمن النتائج لكل عميل." يتطلب التحقق الحقيقي أدلة قوية وشاملة وقابلة للتطبيق للغاية. قد تكون المصادر الرسمية قوية بالنسبة لادعاءات C0 وبعض ادعاءات C1. وحدها، قد لا تكون كافية لمستويات C3–C5.

26. قياسان منفصلان للدقة

يجب الإبلاغ عن قياسين رئيسيين بشكل منفصل ضمن GEO-1000.

26.1. دقة التمثيل الرسمي

يجيب على السؤال التالي:

> إلى أي مدى ينقل استجابة الذكاء الاصطناعي التمثيل الرسمي والرسمى للإصدار للكيان بدقة وبشكل كامل، مع الحفاظ على حالة المصدر؟

يقيس هذا:

- الهوية الرسمية،
- نطاق الخدمة الرسمي،
- السعر المنشور،
- الحدود الرسمية،
- طبيعة المصدر الرسمي

يقيم مجالاتهم.

26.2. دقة حقيقة الكيان التي تم التحقق منها

يجيب على السؤال التالي:

> إلى أي مدى يعكس رد الذكاء الاصطناعي الواقع الحالي للكيان، مدعومًا بالأدلة المتاحة، بدقة وشمول، مع الحفاظ على الحدود؟

يقيس هذا:

- الادعاءات الواقعية المدعومة،
- الادعاءات المتضاربة،
- المعلومات القديمة،
- العلاقات غير الصحيحة،
- القصور الواقعي،
- التحسينات المعرفية

يتم تقييمها.

27. عجز نتيجتين عن تعويض بعضهما البعض

لا يمكن للإخلاص الرسمي العالي أن يعوّض عن الدقة المُحقَّقة المنخفضة. مثال: ينشر الموقع الرسمي ادعاءً خاطئًا. يكرر الذكاء الاصطناعي الادعاء بدقة. قد يكون الإخلاص الرسمي عاليًا. الدقة المُحقَّقة منخفضة. بالمثل، لا تؤدي الدقة المُحقَّقة العالية إلى إلغاء تحريف التمثيل الرسمي تمامًا. مثال: يستخدم الذكاء الاصطناعي مصدرًا خارجيًا محدثًا بشكل صحيح. يخلط بين وصف المنتج والخدمة لشركة واحدة مع أخرى. بعض الحقائق الخارجية في الرد قد تكون صحيحة. الإخلاص في الهوية الرسمية ضعيف. لذلك، يجب عرض النتيجتين بشكل منفصل.

28. حالات الاستجابة ذات الطبقتين

يمكن أن تتلقى استجابة الذكاء الاصطناعي واحدة من هذه النتائج:

تمرير مزدوج

إنه متسق ماديًا مع التمثيل الرسمي والواقع المُحقَّق على حد سواء.

رَسْمِي-نجاح / مُحقَّق-فشل

يحمل السجل الرسمي بأمانة. ومع ذلك، ينقل ادعاءً غير مدعوم أو قديم أو متناقض للسجل الرسمي كما لو كان صحيحًا.

فشل رسمي / اجتاز التحقق

لا يتوافق تمامًا مع الموقع الرسمي. ومع ذلك، ينقل الحقيقة بشكل صحيح مدعومة بأدلة حالية وأقوى. يُفضل توضيح الفرق في الرد.

فشل مزدوج

يعطل مادياً كل من السجل الرسمي والواقع المؤكد.

اجتاز رسمي / الواقع المؤكد غير معروف

تم نقل البيان الرسمي بدقة. لم يتم بعد تحديد حالة الحقيقة المؤكدة للادعاء.

تضارب-المرجع

السجلات المرجعية متناقضة في ذاتها. لا يمكن تقييم الاستجابة بشكل قاطع.

عدم-اليقين-المناسب

لقد شرح الذكاء الاصطناعي بشكل صحيح عدم اليقين المرجعي وتجنب إصدار حكم قاطع. في بعض الحالات، قد يكون هذا هو الجواب الأكثر دقة.

29. عرض APPLE.COM الاصطناعي

عرض المنهجية التركيبية / الادعاءات التالية، الردود، والدرجات كلها خيالية تمامًا. إنها لا تمثل معلومات حقيقية، أو أداء، أو ردود أي منتج ذكاء اصطناعي حقيقي من شركة .Apple Inc سجل التمثيل الرسمي التركيبي: O-:01 تعرف الشركة نفسها على أنها علامة تجارية تكنولوجية استهلاكية عالمية. O-:02 تنشر الشركة بعض فئات المنتجات والخدمات الرقمية. O-:03 تعرف الشركة نفسها بأنها 'الشركة الأكثر ابتكارًا في العالم.' O-:04 يحتوي الموقع الرسمي على شروط متجر محددة للدول. سجل الواقع المثبت التركيبي: V-:01 يتم دعم الهوية المؤسسية

والفئة الرئيسية للأعمال. V-:02 بعض فئات المنتجات مدعومة. V-:03 لا توجد طريقة مقارنة محددة للادعاءات 'أكثر الشركات ابتكارًا في العالم.' V-:04 تختلف ظروف المتجر حسب البلد.

29.1. الاستجابة التركيبية أ

"Apple.com هو أحد الواجهات الرسمية للشركة حيث تقدم الشركة منتجاتها التكنولوجية للمستهلكين وخدماتها الرقمية. تضع الشركة نفسها كواحدة من أكثر الشركات ابتكارًا في العالم؛ يجب أيضًا تقييم هذا الادعاء بالتفوق وفقًا لطريقة المقارنة المستخدمة." التقييم: تم الحفاظ على التمثيل الرسمي. يُنسب ادعاء التفوق إلى الشركة. لم يتم رفع الادعاء إلى حقيقة مستقلة. تمت إضافة تحديد النطاق. حالة المرشح:

تمرير مزدوج

29.2. استجابة تركيبية ب

«أبل هي الشركة الأكثر ابتكارًا بلا منازع في العالم.» التقييم: تم تحويل الادعاء التسويقي الرسمي إلى حقيقة مطلقة. لا توجد طريقة للمقارنة. فقدت حالة المصدر. «بلا منازع» هو رفع معرفي إضافي. حالة المرشح:

رَسْمِي-نجاح / مُحقَّق-فشل

29.3. استجابة تركيبية ج

«أبل هي شركة محلية تبيع الهواتف فقط في المتاجر الفعلية.» التقييم: تم تشويه هوية الشركة ونطاقها. التمثيل الرسمي غير صحيح. يتعارض مع الواقع المثبت. حالة المرشح:

فشل مزدوج

29.4. الاستجابة التركيبية د

تنشر الموقع الرسمي للشركة بعض شروط المتجر؛ ومع ذلك، قد تختلف الأسعار وشروط الشراء وفقًا لإصدار البلد. التقييم: تم استخدام السجل الرسمي بشكل صحيح كمصدر. تم الحفاظ على النطاق الجغرافي. لم يُستخدم الموقع الرسمي كمصدر لسعر واحد في جميع أنحاء العالم. حالة المرشح:

تمرير مزدوج

30. حالة أسترون الاصطناعية

حالة اصطناعية — ليست مؤسسة حقيقية

ينشر الموقع الرسمي لشركة تُدعى أسترون الادعاءات التالية:

- "نحن نعمل في 25 دول."
- "نحقق النجاح في 98% من المشاريع."
- نحن منظمة معايير عالمية تم التحقق منها بشكل مستقل.
- جميع مشاريعنا مضمونة.

سجلات التدقيق:

- الدولة التي تم التحقق من الخدمة النشطة فيها: 9
- بيانات النجاح: المشاريع المختارة فقط 40
- التحقق المستقل: لم يتم العثور عليه
- المعيار: المعيار الموصى به المنشور من قبل الشركة نفسها

- الضمان: تعاقدي فقط لعناصر التسليم المحددة

استجابة الذكاء الاصطناعي: 'أسترون هي منظمة معيارية عالمياً، تعمل في دول 25، بمعدل نجاح 98%، تم التحقق منه بشكل مستقل، وتضمن النتائج في جميع مشاريعها.'

30.1. دقة التمثيل الرسمي

الاستجابة وفية إلى حد كبير لنص الموقع.

30.2. دقة الواقع المؤكد

الاستجابة:

- يوسع نطاق البلد،
- يجعل البيانات المختارة نجاحًا عامًا،
- يعرض المعيار الخاص بالكيان كمعيار مستقل،

يحوّل ضمان التسليم إلى ضمان النتيجة. إنه فشل جسيم من حيث الواقع المؤكد.

30.3. كيف يمكن أن تكون الاستجابة الصحيحة للذكاء الاصطناعي؟

"تذكر Asteron على موقعها الخاص أنها تقدم خدمات في دول 25، ولديها معدل نجاح 98%، وتنشر معياراً. في سجلات التدقيق الحالية، تم التحقق من الخدمة النشطة في تسع دول؛ لوحظ أن معدل 98% مبني على مشاريع 40 مختارة؛ لم يمكن تأكيد التحقق المستقل. الضمان محدود لعناصر التسليم المحددة." هذا الرد:

- لا يخفي بيان الشركة،
- يحافظ على وضع البيان،
- يميز الواقع المؤكد,

يجعل القيد مرئيًا.

31. حالة تركيبية تحتوي على خلل في المرجع.

الحالة الاصطناعية

الصفحة الرئيسية الرسمية: "السعر الابتدائي 2,000 يورو." ملف PDF للسعر الرسمي: "السعر الابتدائي 3,000 يورو." الدفع: "4,000 يورو." البيانات المنظمة: "1,500 يورو." استجابة الذكاء الاصطناعي: "السعر الابتدائي هو 3,000 يورو." أي سجل اختاره الذكاء الاصطناعي؟ ملف PDF. هل هو خطأ؟ ربما. هل هو صحيح؟ ربما. السجل المرجعي غير متسق. في هذه الحالة:

- فشل الذكاء الاصطناعي،
- تضارب التمثيل الرسمي،
- نقص المصدر القانوني

هي نتائج منفصلة. التقييم الصحيح:

تضارب-المرجع

ممكن. لا يمكن للكيان توقع الدقة الكاملة للأسعار دون توحيد سجلات الأسعار الخاصة به.

32. أولوية المصدر في طبقة التمثيل الرسمي

عندما يتم العثور على سجلات رسمية متعددة، يجب وضع قاعدة أولوية لكل حقل ادعاء. مثال على أولوية المرشحين:

- سجل العقد أو المعاملة الحالي
- سجل المنتج/السعر المحدث الرسمي
- البيانات المهيكلة المصرح بها
- الصفحة الرسمية الحالية
- ملف PDF الرسمي الحالي
- السجل الصحفي أو المدونة التاريخية

هذا الترتيب ليس عالمياً لجميع الادعاءات. على سبيل المثال، قد تكون أرشيفات الصحف قيمة لتاريخ الشركات. السجل القديم للمدونة ليس مصدراً موثوقاً للسعر. يجب تحديد الأولوية وفقاً لنوع الادعاءات.

33. تعارض المصدر في الواقع المؤكد

قد تتناقض مصدراان مستقلان مع بعضهما البعض. في هذه الحالة:

- الأكثر استقلالية منهما،
- الأحدث منهما،
- الأقرب إلى الموضوع،
- الأكثر وضوحاً في منهجه،
- المستخدم للبيانات الأوسع

لا يفوز تلقائيًا. يجب تقييم جميع المعايير معًا. مثال: يمكن أن يكون منشور مدونة جديد أكثر حداثة من سجل رسمي قديم. ومع ذلك، قد لا تتمتع المدونة بالسلطة الكافية للملكية القانونية. يجب أن يكون حل تعارض المصادر حساسًا لنوع الادعاء.

34. الحفظ الإلزامي لـ الأدلة المضادة

لا يمكن أن يتضمن السجل المؤكد للحقائق سجلات تدعم الادعاء فقط. يجب أيضًا الحفاظ على ما يلي:

- السجلات التي تتناقض مع الادعاء
- القضايا غير الناجحة
- النتائج المسحوبة
- الشهادات منتهية الصلاحية
- النتائج السلبية للمستخدم
- القرارات التنظيمية
- التصحيحات والأخطاء
- التناقض بين المصادر

دعم الادعاء لا يسمح برفض الأدلة المضادة.

المادة 35

ليست كل الفروق المؤكدة رسميًا لها نفس مستوى الأهمية. الفرق المادي هو الفرق الذي قد يغير قرار المستخدم أو النظام. أمثلة: فرق بسيط في تهجئة اسم الشركة قد لا يكون ماديًا. كتابة نوع الشركة بشكل خاطئ قد تكون مادية. غياب لون المنتج قد لا يكون ماديًا. ذكر بلد الخدمة بشكل غير صحيح يكون ماديًا. فرق في شكل التاريخ قد لا يكون ماديًا. تاريخ صلاحية الترخيص غير الصحيح يكون ماديًا. تغيّر نغمة التسويق قد لا يكون ماديًا. إزالة حد الضمان تكون مادية. يجب تقييم المادية بناءً على تأثيرها على:

- قرار المستخدم،
- المسؤولية القانونية،
- السعر،
- أمني،
- السلطة،
- النطاق،
- الأدلة.
- الوقت

يجب تقييم هذه الفئات بناءً على تأثيرها على قرار المستخدم، المسؤولية القانونية، السعر، الأمان، السلطة، النطاق، الأدلة والوقت.

36. الأحكام الإلزامية المعيارية

CH04-N01

يجب تقييم كل رد مادي من الذكاء الاصطناعي بشكل منفصل مقابل سجل التمثيل الرسمي وسجل الواقع الكيان المؤكد.

CH04-N02

لا يمكن اعتبار دقة التمثيل الرسمي نفس نتيجة الدقة المادية.

CH04-N03

لا يُعتبر الادعاء على موقع الكيان الخاص تحققًا مستقلاً من دون وجود دليل إضافي مناسب لنوع الادعاء.

CH04-N04

لا يمكن رفض الدليل من الطرف الأول لمجرد كونه من الطرف الأول؛ يجب تقييم ملاءمته لنوع الادعاء والطريقة وتضارب المصالح.

CH04-N05

لا يمكن قبول المصدر المستقل على أنه صحيح أو موثوق لمجرد كونه مستقلاً.

CH04-N06

لا يمكن رفع رأي الشركة أو ادعاء التسويق أو التقييم الذاتي إلى مستوى الحقيقة المستقلة في رد الذكاء الاصطناعي.

CH04-N07

يجب الحفاظ على نوع المصدر والحالة المعرفية في رد الذكاء الاصطناعي إلى الحد الذي يكون ذا صلة.

CH04-N08

لا يمكن اعتبار استجابة الذكاء الاصطناعي التي تتناقض مع السجل الرسمي فشلًا تلقائيًا إذا كانت مدعومة بأدلة أقوى وأكثر حداثة.

CH04-N09

لا يمكن اعتبار استجابة الذكاء الاصطناعي المخلصة للسجل الرسمي ناجحة من حيث الدقة العامة إذا كان الادعاء الرسمي يتناقض مع الحقيقة المثبتة.

CH04-N10

إذا تعارضت السجلات الرسمية والمحققة ماديًا، يجب أن تكون النتيجة CONFLICT_REFERENCE، أو UNRESOLVED، أو تحمل حالة عدم يقين مناسبة.

CH04-N11

لا يمكن تحويل عيب في المرجع بهدوء إلى خطأ في منتج الذكاء الاصطناعي.

CH04-N12

إذا تعامل رد الذكاء الاصطناعي مع ادعاء طرف أول غير مؤكد كحقيقة، فيجب السعي للحصول على الاقتباس والتقييد المناسب.

CH04-N13

لا يمكن ترك الحذوفات الرسمية المادية لتبدو ببساطة كما لو أن الجملة المنشورة صحيحة.

CH04-N14

يجب أيضًا تسجيل الأخطاء التي تفيد الشركة كأخطاء في طبقة الواقع المحقق.

CH04-N15

يجب توثيق سجل التمثيل الرسمي وسجل واقع الكيان المحقق وإصدارهما قبل بدء القياس.

CH04-N16

إذا حدث تغيير في المرجع المادي خلال موجة القياس، يجب تقييم الملاحظات السابقة واللاحقة في نسخ منفصلة.

CH04-N17

لا يمكن إزالة الأدلة المضادة والسجلات السلبية من ملف الواقع المؤكد لمجرد أنها سلبية.

CH04-N18

لا تُعتبر نتائج غير محلولة، أو غير معروفة، أو الأدلة الغير كافية تحققًا إيجابيًا.

CH04-N19

كلما زادت قوة الادعاء، يجب أن يزداد أيضًا عبء الإثبات المطلوب.

CH04-N20

لا يمكن للدقة العالية في التمثيل الرسمي تعويض انخفاض دقة واقع الكيان المحقق.

CH04-N21

يجب أن يكون هناك صاحب مسؤول من الإنسان أو المؤسسة نتيجة للتقييم ذو الطبقتين.

37. أشكال الفشل

CH04-F01— اعتبار الموقع الإلكتروني واقعًا مطلقًا

يتم قبول جميع الادعاءات المنشورة من قبل المؤسسة كحقائق مُحققة.

CH04-F02— اعتبار السجلات الرسمية عديمة القيمة

يتم تجاهل السجلات المؤسسية المباشرة مثل السعر أو السياسة أو تغطية المنتج لأنها من الطرف الأول.

CH04-F03— جعل الإعلان الذاتي دليلًا مستقلًا

يتم استخدام قيادة الشركة أو ادعاء نجاحها كما لو كان تحققًا خارجيًا.

CH04-F04— عبادة المصادر المستقلة

المصادر المستقلة ولكن البعيدة عن الموضوع، القديمة، أو غير المنهجية تعتبر متفوقة على السجلات الرسمية والقريبة.

CH04-F05— الارتفاع المعرفي

تُحوَّل عبارة "تقول الشركة" إلى الحكم "لقد تم إثباته."

CH04-F06— مسح حالة المصدر

يتم تقديم بيانات الطرف الأول، أو المحتوى المدعوم، أو البيانات الصحفية على أنها أخبار مستقلة.

CH04-F07— إعادة إنتاج الأخطاء الرسمية بأمانة

الذكاء الاصطناعي يكرر الادعاء الرسمي الخاطئ بشكل صحيح ويعتبر ناجحًا بشكل عام.

CH04-F08— معاقبة التصحيحات الخارجية المحدثة

يعتبر الذكاء الاصطناعي مخطئًا لأنه يغادر الموقع الرسمي القديم ويستخدم الأدلة الحالية القوية.

CH04-F09— اعتبار الصمت الرسمي ممنوعًا

أي معلومات خارجية غير موجودة على الموقع الرسمي تعتبر تلقائيًا غير صحيحة.

CH04-F10— تجاهل النقص المادي

حتى إذا كانت الجملة الرسمية صحيحة، لا يعتبر الحد اللازم.

CH04-F11— النظر في خطأ المرجع كعيب في النموذج

يتم لوم الذكاء الاصطناعي وحده على سجلات الأسعار أو الهوية المتناقضة.

CH04-F12— إخفاء تعارض المرجع

يختار المحكمون المصدر الذي يريدونه من بين السجلات المتناقضة بعد النتيجة.

CH04-F13— ترتيب تسلسلي للمصادر المستقلة بغض النظر عن نوع الطلب

في أي حال، يُفترض أن المصدر الرسمي أو المستقل هو الأفضل.

CH04-F14— حذف الأدلة المعارضة

يتم إزالة السجلات التي تضعف الادعاء الإيجابية من ملف التحقق.

CH04-F15— اعتبار السجلات الرسمية القديمة على أنها حالية

الأرشيفات، ملفات PDF القديمة، أو البيانات الصحفية تُعتبر المصدر القانوني الحالي.

CH04-F16— احتساب الادعاءات غير المؤكدة كناجحة

نظرًا لأن الكذب لم يُثبت، يُعتبر الادعاء مؤكَّدًا.

CH04-F17— المبالغة في تقدير قوة الادعاء استنادًا إلى الأدلة

يتم تعميم بيانات تجريبية محدودة أو محلية على نتائج عالمية.

CH04-F18— درجة جغرافية ذات طبقة واحدة

يتم دمج الدقة المُوثَّقة بالأمانة الرسمية بشكل غير مرئي ضمن نتيجة واحدة.

CH04-F19— الحفاظ على التحيز المواتي
لا يتم عمل أي تصحيح عندما يبالغ المؤسسة الذكية.
CH04-F20— عد سجل الواقع كقرار تأسيسي
يتم تحويل بيان NobleJackal أو Kaan MURAZ إلى واقع مُثبت دون أي دليل خارجي.

38. إجراء التدقيق

الخطوة 1— استخراج الادعاءات الذكية المادية

الاستجابة:

- الادعاءات الذرية،
- العلاقات،
- نطاق ومناطق زمنية

يتم فصلها.

الخطوة 2— تحديد موضوع كل ادعاء

الادعاء تتعلق بما:

- شركة
- علامة تجارية،
- المنتج,
- شخص،
- كيان محلي

؟

الخطوة 3— العثور على سجل التمثيل الرسمي

هل وُجد الادعاء على الواجهات الرسمية؟ في أي مصدر؟ في أي نسخة؟ في أي تاريخ؟ ضمن أي نطاق؟

الخطوة 4— تحديد الوضع الرسمي

ادعاء:

- قَنَوِيّ
- تاريخي،
- منسحب
- مسودة
- متعارض،
- غير معروف

أي من أوضاعهم هو؟

الخطوة 5— تصنيف نوع الادعاء

ادعاء:

- هوية

- السعر،
- النطاق،
- رخصة،
- أداء
- القيادة
- شراكة،
- نتيجة المستخدم

أي من أنواعه هو؟

الخطوة —**6** تحديد النوع المناسب من الأدلة

أي مصدر هو بالفعل موثوق لادعاء معين؟ لا يتم استخدام ترتيب مصدر عالمي واحد.

الخطوة —**7** جمع التحقق و الأدلة المضادة

يتم تقييم السجلات الداعمة، السجلات المتضاربة، الالتزام بالمواعيد، النطاق، الطريقة، والاستقلالية.

الخطوة —**8** تعيين حالة التحقق

يتم تعيين الحالات من VR-0 إلى VR-5، VR-T، VR-U، VR-C، أو VR-S حسب الاقتضاء.

الخطوة —**9** مراجعة حماية حالة المصدر

الذكاء الاصطناعي: هل حافظ على الفارق بين "تقول الشركة" و"مثبت"؟ هل رفع الإعلان الذاتي إلى درجة الواقع المستقل؟

الخطوة —**10** فحص النواقص المادية

في الاستجابة أو التمثيل الرسمي، الضروري لاتخاذ قرار المستخدم:

- الحدود،
- الوقت،
- الجغرافيا،
- رخصة،
- طريقة البيانات

مفقود؟

الخطوة —**11** فصل عيب المرجع

إذا كان السجل الرسمي أو المعتمد معيبًا في حد ذاته، يتم إنشاء نتيجة منفصلة عن خطأ الذكاء الاصطناعي.

الخطوة —**12** تقديم نتيجة ذات طبقتين

الاستجابة:

تمرير مزدوج

رَسْمِي-نجاح / مُحقَّق-فشل

فشل رسمي / اجتاز التحقق

فشل مزدوج

تضارب-المرجع

عدم-اليقين-المناسب

يتم حفظه بالحالة المناسبة.

39. الأدلة المطلوبة

سجل الكيان القانوني السجل الرسمي للتمثيل سجل الكيان الموثق سجل الواقع الرسمي للصفحات الإلكترونية الرسمية الملفات والوثائق الرسمية البيانات المنظمة السجلات القانونية JSON أو سجلات API السجلات القانونية والتنظيمية التحقق من الرخص والشهادات سجلات المنتجات والأسعار وعمليات الدفع سجلات العقود والتشغيل التحقق من العملاء والشركاء مجموعات بيانات الأداء المنهجية والمقاييس المراجعات المستقلة الرعاة وتضارب المصالح الأدلة المضادة النتائج السلبية سجلات التصحيحات والسحب قواعد أسبقية المصادر الرسمية حالات التحقق قرارات المحكم سجلات الاستئناف تاريخ القياس الإصدارات المرجعية الشخص أو المؤسسة المسؤولة

40. قائمة تدقيق التدقيق

هل هناك سجل تمثيل رسمي؟ هل هناك سجل واقع كيان مُحقق؟ هل تم إصدار نسختين لنفس تاريخ القياس؟ هل موضوع كل ادعاء ذكاء اصطناعي صحيح؟ هل تم نشر الادعاء رسميًا؟ هل المصدر الرسمي حديث وقانوني؟ هل تتعارض السجلات الرسمية مع بعضها البعض؟ هل تم استخدام مصدر الأدلة المناسب لنوع الادعاء؟ هل تم رفض مصدر من الطرف الأول تلقائيًا؟ هل تم التعامل مع بيان من الطرف الأول تلقائيًا كدليل مستقل؟ هل المصدر المستقل ذو صلة فعلية؟ هل تم الحفاظ على نطاق الادعاء والوقت؟ هل يحتوي الملف على الأدلة المضادة؟ هل تم ترقية إعلان ذاتي إلى واقع مُحقق؟ هل تم نقل وجهة نظر الشركة مع الحفاظ على حالة المصدر الخاصة بها؟ هل هناك حد مادي أو استثناء مفقود؟ هل تم تسجيل خطأ لصالح الكيان على هذا النحو؟

هل تم فصل العيب المرجعي عن خطأ الذكاء الاصطناعي؟ هل تم اعتبار الحقول UNKNOWN و UNRESOLVED إيجابية؟ هل تخفي الأمانة الرسمية العالية دقة منخفضة تم التحقق منها؟ هل يمكن تفسير انحراف الذكاء الاصطناعي عن السجل الرسمي بأدلة قوية؟ هل تغير السجل المرجعي أثناء القياس؟ هل الملاحظات القديمة والجديدة في إصدارات منفصلة؟ هل يظهر النتيجة ذات الطبقتين في السجل العام؟ هل يعرف المالك المسؤول عن النتيجة؟

41. الاعتراضات والردود

اعتراض **1**— "تعرف الشركة على أكثر المعلومات دقة عن نفسها."

قد يكون هذا صحيحًا في نواحٍ عديدة. الشركة:

- منتجه الخاص,
- سعره،

- سياسة الدعم الخاصة بها،
- تشغيله

يعرف ذلك جيدًا. ومع ذلك، فإن الشركة هي أيضًا المالك التجاري للادعاء. الجمل التي تتطلب مقارنة خارجية، مثل "الأفضل"، "القائد"، "الأكثر موثوقية"، و"المثبتة"، لا يمكن التحقق منها بالإعلان الذاتي وحده.

اعتراض 2— "إذا لم تحسب موقعك الخاص كدليل، كيف سنعرف السعر؟"

صفحة السعر الرسمية هي المصدر الرئيسي للسعر الذي تقدمه المؤسسة للجمهور. ومع ذلك:

- الدفع عند الخروج،
- الضرائب،
- الرسوم الإضافية،
- العقد

إذا كان هناك تضارب، يتم تقييم العرض الفعلي بشكل منفصل. لا يمكن تطبيق العبارة "الموقع الخاص ليس دليلًا" بنفس الطريقة على جميع أنواع الادعاءات. الحكم الصحيح هو:

> يمكن لموقع الشركة الخاص أن يكون مصدرًا رئيسيًا للبيانات تحت سيطرتها؛ لكنه ليس تحققًا مستقلاً تلقائيًا للادعاءات المتعلقة بالأداء والتفوق الخارجي.

اعتراض 3— "إذا كان الذكاء الاصطناعي يعكس الموقع الرسمي، أليس قد أتم مهمته؟"

إذا كان الهدف الوحيد هو: «ماذا يقول الموقع؟» فإن الإجابة نعم. ومع ذلك، إذا سأل المستخدم: «ماذا تفعل الشركة فعليًا؟» فإن مجرد تكرار ما يقوله الموقع قد لا يكون كافيًا. نوع المهمة يؤثر على طبقة القياس.

اعتراض 4— "المصادر الخارجية يمكن أن تكون خاطئة أيضًا."

هذا صحيح. لذلك، الاستقلال ليس دقة تلقائية. يتم تقييم المصدر من حيث:

- السلطة،
- القرب
- عملة
- طريقة
- نطاق

من حيث الصيانة.

اعتراض 5— "من سيحدد السجل المثبت للواقع؟"

لا يمكن للشركة وحدها تحديد ذلك. ولا يمكن لـ NOMOS وحده تحديده أيضًا. العملية:

- حامل الأدلة،
- المحكم،
- الخبير المناسب،
- الاستئناف،
- سجل النسخة،
- الشخص المسؤول

مطلوب. هناك حاجة لمراجعة مستقلة أكثر للادعاءات عالية المخاطر والتي بها تضارب مصالح.

اعتراض 6— "إذا كان يمكن دائمًا تغيير سجل الواقع، ألا يعني ذلك أن النتيجة ستكون بلا معنى؟"

لا. النتيجة:

- تاريخ،
- ذات معنى مع
- نطاق

النسخة. كونها عرضة للتغيير لا يعني أنها لا يمكن قياسها. يعني أنه لا يمكن عرضها قبل الأوان.

اعتراض 7— "لماذا يستخدم الذكاء الاصطناعي معلومات سلبية غير موجودة على الموقع الرسمي؟"

المعلومات:

- أن يُجبر على الانتماء إلى فئات

صحيح،

- حالي،
- ذات صلة،
- مالية لقرار المستخدم

يمكن استخدامه. ومع ذلك، يجب الحفاظ على المصدر والسياق. صمت الموقع الرسمي لا يزيل الحقيقة السلبية.

الاعتراض 8— 'إذا أبلغ الذكاء الاصطناعي عن ادعاء الشركة بأنه "تقول الشركة"، فهذا لا يضيف قيمة للمستخدم.'

على العكس من ذلك، فإنه يعرض حالة الادعاء بشكل صحيح. المستخدم:

- رأي المؤسسة،
- النتيجة المثبتة،
- الغموض المفتوح

يمكن التمييز. المظهر المؤكد ليس هو نفسه المظهر الصحيح.

اعتراض 9— "لماذا يجب أن نصحح الأخطاء البسيطة التي تصب في مصلحة الشركة؟"

بسبب التوقع الخاطئ:

- العميل الخطأ،
- فشل التسليم،
- الإرجاع،
- الشكوى،
- فقدان الثقة

يمكن أن ينتج. الميزة هي وضوح قصير المدى؛ على المدى الطويل هو واجب التمثيل.

اعتراض 10— "ألا تجعل درجتان منفصلتان الاتصال العام أكثر تعقيدًا؟"

إخفاء كل هذه التعقيدات يخلق خطراً أكبر. قد يُشير بطاقة النتائج المختصرة بدلاً من ذلك إلى: دقة التمثيل الرسمي — عالية / دقة واقع الكيان المؤكد — متوسطة / الفجوات الرئيسية — ادعاءات القيادة والتغطية القطرية غير المدعومة. هذا أوضح من رقم قد يضلل.

1. سجل التمثيل الرسمي؛ و

2. سجل الواقع للكيان الموثق والمقيد بالأدلة.

يجب عدم التعامل مع الأمانة الرسمية والدقة المؤكدة على أنهما نفس النتيجة. قد يقوم رد الذكاء الاصطناعي بإعادة إنتاج ادعاءات رسمي بأمانة بينما يظل غير مدعوم بالحقائق، أو متناقضًا، أو قديمًا، أو واسع النطاق بشكل مفرط، أو مضلل جوهريًا. قد يختلف رد الذكاء الاصطناعي عن السجل الرسمي دون أن يُعد فشلًا عندما يكون السجل الرسمي قديمًا، أو ناقصًا، أو متناقضًا، أو يتعارض مع أدلة أقوى وأكثر ملاءمة. يجب أن تحتفظ الادعاءات الذاتية، والبيانات من الطرف الأول، والأدلة المستقلة، والمحتوى الممول، والسجلات الرسمية، والسجلات المعاملاتية، وبيانات الطرف الثالث بمصدرها وحالة الأدلة الخاصة بها. يجب تسجيل تعارضات المراجع بشكل منفصل عن إخفاقات ردود الذكاء الاصطناعي. لا يجوز تحويل الأدلة المجهولة أو غير المحلولة إلى حقيقة مؤكدة.

يجب تقييم كل ادعاء من نظام توليدي مادي عن الكيان الذي تم تدقيقه مقابل طبقتين مرجعيتين منفصلتين الإصدار: سجل التمثيل الرسمي وسجل الواقع الكياني المثبت والمقيد بالأدلة. لا يجوز التعامل مع الوفاء الرسمي والدقة المثبتة كنتيجة واحدة. قد يقوم منتج ذكاء اصطناعي بإعادة إنتاج ادعاء رسمي بدقة بينما ينشئ تمثيلاً غير مدعوم ماديًا، أو متناقضًا، أو قديمًا، أو عامًا بشكل مبالغ فيه. عندما يكون السجل الرسمي قديمًا أو غير مكتمل أو متناقضًا، فإن استجابة مختلفة مستندة إلى أدلة أقوى ليست تلقائيًا خطأً. يجب الحفاظ على نوع المصدر وحالة الأدلة، ويجب تسجيل تعارضات المراجع بشكل منفصل عن أخطاء الذكاء الاصطناعي. لا يجوز أن يعوّض الوفاء العالي بالتمثيل الرسمي عن انخفاض دقة الواقع الكياني المثبت.

الحكم العام للفصل 44

لدى الشركة الحق في التحدث عن نفسها. يمكنها شرح من هي. يمكنها وصف ما تقدمه. يمكنها نشر أسعارها. يمكنها تحديد حدودها. يمكنها نقل رؤيتها. ومع ذلك، ليس كل جملة تقولها الكيان عن نفسها لها نفس الوضع المعرفي. "سعر حزمة خدمتنا هو 2,000 يورو." و "نحن أفضل شركة في العالم." ليستا نفس نوع الادعاء. الجملة الأولى قد تكون تصريحًا تجاريًا تتحكم فيه الشركة. الجملة الثانية تتطلب مقارنة خارجية. إذا نقل منتج ذكاء اصطناعي كلا الجملتين بنفس التأكيد، فإنه يقلل من مكانة المصدر. يمكن لمنتج الذكاء الاصطناعي أن يكرر موقع الشركة تمامًا. ولكن الموقع:

- قديم،
- متعارض،
- غير مثبت،
- يتجاوز النطاق،
- يخفي الحدود المادية

ثم الوفاء بذلك السجل ليس جيدًا GEO. وعلى العكس، قد ينحرف منتج الذكاء الاصطناعي عن الموقع الرسمي ومع ذلك يكون صحيحًا عندما يستند جوابه إلى دليل أقوى وحديث ومناسب. قد يكون الانحراف تصحيحًا بدلًا من خطأ. لذلك فإن القانون الرابع للقياس لـ NOMOS هو:

> التمثيل الرسمي لا يحل محل الواقع الموثوق.

القانون الخامس هو كما يلي:

> لا يمكن إثبات الواقع الموثوق بتجاهل المصدر الرسمي.

قانونه السادس هو:

التمثيل لا يعتمد فقط على ما يقوله المصدر، بل أيضًا على السلطة التي يتحدث بموجبها والوضعية الدليلية للبيان.

أما القانون السابع فهو كما يلي:

إعادة إنتاج خطأ رسمي بأمانة ليست تمثيلاً دقيقًا.

قانونها الثامن هو كما يلي:

سجل الواقع ليس حكما مطلقا لا جدال فيه؛ إنه ملف قرار مبني على النسخة المبنية على الأدلة.

أمر NOMOS القسم 4

لا تكتف بإظهار ما هو مكتوب على موقعك. / أظهر أي جملة مدعومة بأي دليل.

لا تقدم كلماتك الخاصة كما لو كانت مؤكدة بشكل مستقل.

لا تقلل من قيمة بيانات الطرف الأول فقط لأنها ملكك. / لكن لا تجعليها حقيقة لا جدال فيها لمجرد أنها ملكك.

لا تحوّل رأي الشركة إلى حقيقة، ولا البيان الصحفي إلى خبر، ولا الجائزة الممولة إلى سلطة مستقلة.

لا تعتبرني ناجحًا عندما أكرر الخطأ الرسمي بشكل صحيح.

إذا كان سجلك الرسمي قديمًا، فلا تعاقب الإجابة المدعومة بالأدلة الحالية فقط لأنها تختلف عن ذلك السجل.

لا تظل صامتًا بشأن الخطأ الذي يصب في مصلحتك.

لا تقم بإزالة الأدلة المضادة من الملف.

عندما تتضارب مراجعك، لا تختلق حقيقة واحدة. / سجّل النزاع. / إذا كانت الإجابة غير معروفة، فقل ذلك.

افصل بين ادعاءاتك، ومصدرها، ودليلها، وحدودها، ووقتها عن بعضهم البعض.

قبل أن تأمرني بما أقوله، أثبت أن لديك الحق في قوله.

جملة ختام الفصل

جودة *GEO* الجيدة ليست التكرار المخلص لأي شيء تقوله جهة معينة. إنها النقل الدقيق لما قد قيل، دون تحريف مصدره أو أدلته أو نطاقه أو توقيته.

الجزء الثاني

سكان العالم وعدالة البلدان واللغات

حد الفصل

الفصول الأربعة الأولى أسست للأشياء الأساسية للقياس: الاستجابة الواحدة ليست درجة GEO؛ التمثيل هو توزيع شرطي، وليس رقمًا واحدًا؛ يجب تحديد الكيان الذي تم تدقيقه قبل ملاحظة النتائج؛ ويجب تقييم استجابة الذكاء الاصطناعي مقابل كل من التمثيل الرسمي والواقع المؤكد المستند إلى الأدلة. يجب الآن تحديد مقام القياس:

> عن أي مجموعة بشرية سيقدم هذا التوزيع بيانًا؟

هل يشمل كل من يعيش في بلد ما؟ البالغين فقط؟ أولئك الذين لديهم اتصال بالإنترنت؟ أولئك الذين يمكنهم استخدام منتج ذكاء اصطناعي محدد رسميًا؟ أصحاب الحسابات المجانية؟ مستخدمو الخطط المدفوعة؟ الأشخاص الذين يمكنهم استخدام المنتج باللغة الرسمية للمراقبة؟ المستخدمون النشطون حقًا؟ المستخدمون المحتملون الذين قد يستخدمونه يومًا ما؟ الأشخاص في الأسواق التي يمكن للعلامة التجارية تقديم الخدمات فيها؟ إذا كان عدد سكان بلد ما كبيرًا جدًا ولكن المنتج المعني للذكاء الاصطناعي غير متاح رسميًا في ذلك البلد، فكيف سيتم إدراج ذلك البلد في لوحة المنتج الرئيسية؟ إذا كان نصيب السكان في بلد صغير أقل بكثير من شخص واحد في عينة من 1,000، فهل سيكون ذلك البلد غير مرئي تمامًا؟

إذا كان شخص واحد يعرف ثلاث لغات، هل يجب أن يُحتسب هذا الشخص كثلاثة أشخاص؟ إلى أي فئة سكانية ينتمي المستخدم إذا كان يمكنه الوصول إلى واجهة تطبيق المنتج على الهاتف المحمول ولكن ليس على واجهة الويب؟ حتى يتم الإجابة على هذه الأسئلة، فإن '1,000 مستخدمين عشوائيين لكل منتج ذكاء اصطناعي' لا يحدد فئة سكانية علمية؛ إنه يذكر فقط عدد الرؤوس. هذا الفصل لا يتجاهل عدد سكان البلد. إنه يخصص له دوره الصحيح:

> عدد سكان البلد هو الأساس الأولي لأخذ العينات؛ وليس هو عدد المستخدمين النهائيين المؤهلين.

في هذا القسم، يطبق نظام التدقيق مباشرةً المتطلبات التي تم تحديدها مسبقًا بالنسبة للبلد، اللغة، التاريخ، المنتج، وسجل القياس على تعريف العينة. لأن طلب الأدلة، الحدود، السياق، والزمن يتطلب أيضًا تحديد الأشخاص المشار إليهم بنفس الوضوح. هذا القسم لم يقم بعد بـ:

- لا يحدد بعد كيفية توزيع عينة الأشخاص 1,000 عبر البلدان،
- أي خوارزمية ستُستخدم لتقريب الكسور إلى أعداد صحيحة،
- من أي المشاركين في اللوحة سيتم الاختيار،
- كيفية حساب أوزان العينات،
- أو أي طريقة التباين ستُستخدم لإنتاج فواصل الثقة

سيتم ترتيب هذه في الأقسام التالية. مهمة القسم 5 أكثر أساسية:

> لكل منتج ذكاء اصطناعي ولكل سؤال قياس، لتحديد أي سكان بشري يمكن تعميم النتائج عليهم.

التحدي NOMOS

تفتح خريطة العالم. تجمع سكان الدول. ثم ترغب في توزيع 1,000 شخصًا وفقًا للسكان. الدول الكبيرة ستحصل على العديد من المشاركين. الدول الصغيرة ستحصل على حصة واحدة أو قريبة من الصفر. جوهر هذه الفكرة صحيح:

> يجب ألا يكون الشخص في سكان العالم أكثر تأثيرًا بمئات المرات من شخص في دولة كبيرة لمجرد أنه يعيش في دولة صغيرة.

ولكن الآن طبق نفس الطريقة على منتجين ذكاء اصطناعي مختلفين. المنتج الأول:

- يمكن استخدامه رسميًا في بعض الدول،
- محدود في بعض الدول.

لا يُقدَّم على الإطلاق في بعض الدول. تغطية الدولة للمنتج الثاني مختلفة. ماذا يحدث إذا قسمت سكان العالم بنفس الطريقة بين المنتجين؟ في الدول التي لا يتوفر فيها المنتج الأول، يمكنك طلب قياسات من الناس. المشاركون:

- قد يتم تشجيعهم على تقديم معلومات مضللة عن مواقعهم،
- استخدام طرق وصول محظورة أو غير مدعومة،
- فتح حساب في دولة أخرى،
- تجاوز الشروط الأمنية وشروط الاستخدام

هذا لا يقيس تجربة المستخدم الحقيقية. إنه يولد تجربة وصول اصطناعية. الآن اعتبر مشكلة أخرى. في دولة معينة، يبلغ عدد السكان كبيراً جداً. ومع ذلك:

- جزء كبير من السكان لا يستوفي الحد الأدنى للعمر المطلوب للمنتج.
- ليس لدى الجميع وصول موثوق إلى الإنترنت،
- يمكن استخدام المنتج فقط على أجهزة معينة،
- لا يمكن للجميع استخدام لغة التدقيق.

الخطة المحددة متاحة فقط للمستخدمين الذين لديهم وسيلة دفع. إذا جعلت إجمالي السكان هو المقام مباشرة، فإنك تشمل أشخاصًا لا يمكن مراقبتهم فعليًا ولا يمكنهم استخدام المنتج ضمن السكان المستهدفين. الآن تخيل دولة صغيرة. الحصة النظرية في عينة 1,000 وفقًا للأوزان السكانية: 0.03 شخص.

ثلث شخص لا يمكن اختياره. إذا أعطيت شخصًا واحدًا للدولة، يمكنك تمثيل الدولة بشكل مبالغ فيه في اللوحة الرئيسية المُوزونة بالسكان. إذا لم تعطِ أحدًا، فحينها في تلك الدولة:

- الوصول،
- اللغة،
- سلوك المنتج،
- التمثيل المحلي

يمكن أن تظل غير مرئية تمامًا. هذان الهدفان مختلفان:

- لقياس متوسط تجربة المستخدم في سكان العالم
- لمنع تحوّل الدول الصغيرة إلى غير مراقبة تمامًا

لا يمكن للوحة واحدة أن تحقق الهدفين بشكل مثالي. لذلك، يضع NOMOS ثلاثة مبادئ منفصلة:

> وزن السكان يحدد حصة الأشخاص في المتوسط العالمي.
>
> تغطية المراقبين تمنع اختفاء الدول واللغات الصغيرة تمامًا.

العدالة اللغوية وتوفير إمكانية الوصول تقيس بشكل إضافي الفئات الفرعية التي قد يتم تجاهلها في المتوسط السكاني.

المقدمة الأولى لهذا القسم هي:

إجمالي سكان الدولة ليس هو عدد المستخدمين المؤهلين للذكاء الاصطناعي في تلك الدولة.

ينص حكمه الثاني على:

لا يمكن نقل الفئة المستهدفة من مستخدمي منتج الذكاء الاصطناعي تلقائيًا إلى منتج ذكاء اصطناعي آخر.

تنص الفقرة الثالثة منه على:

وزن الشخص في النتيجة الرئيسية ليس هو نفسه حق بلده في الظهور في التدقيق.

تنص المادة الرابعة على:

نقص الوصول إلى المنتج ليس تحريفًا؛ ومع ذلك، فهو قيد وصول عالمي ويجب الإبلاغ عنه أيضًا.

1. غرض الفصل

الغرض من هذا القسم هو تعريف الفئة البشرية المستهدفة لقياس GEO-1000 من حيث المنتج، والسؤال، والبلد، واللغة، والواجهة، والوقت. يقوم هذا القسم بتوحيد التمييزات التالية:

- إجمالي سكان البلد وسكان مستخدمي الذكاء الاصطناعي المناسبين
- الوصول المحتمل والاستخدام الفعلي
- إمكانية الوصول إلى المنتج ودقة التمثيل
- السكان المستهدفون وإطار العينات
- ملاحظة تحليلية صالحة مع المشارك المختار
- الحساب مع الشخص
- الجلسة مع الشخص
- النموذج الأساسي مع المنتج
- سكان الخطط المجانية والمدفوعة
- سكان مستخدمي الويب والجوال
- الاستخدام الذي يتجاوز قيود الوصول مع الوصول الرسمي
- تغطية مراقب الدولة مع وزن سكان الدولة
- عينة عدالة اللغة مع حصة اللغة الموزونة حسب السكان
- التعرض النشط للمستخدم مع السكان المحتملين للمستخدم
- استفسار مجال السوق أو الاختصاص القضائي مع استفسار الهوية العالمية
- المنتج المؤسسي أو التعليمي أو الخاص بالمطور مع المنتج الرئيسي للمستهلك
- سهولة وصول أداة جمع البيانات للمستخدمين ذوي الاحتياجات المتعلقة بإمكانية الوصول
- المستخدمون خارج السكان المستهدفين والمستخدمون الغير مختبرين
- الدول التي يكون المنتج فيها غير قابل للاستخدام والدول التي يتم فيها الحصول على استجابات غير صحيحة في المنتج

في نهاية هذا القسم، يجب أن يكون كل تدقيق GEO-1000 قادرًا على الإجابة صراحة على السؤال التالي:

ما هو كون المستخدمين، المختار من أي أشخاص، الذي يجعل هذه النتيجة يصدر حكماً عنه؟

2. الأحكام المعيارية المركزية

قبل ملاحظة النتائج، يجب على كل تدقيق *GEO-1000* تحديد وإصدار نسخة من مجموعة المستخدمين المستهدفة الخاصة بمنتج الذكاء الاصطناعي الذي يتم تدقيقه، وسؤال القياس، والبلد، واللغة، وواجهة الاستخدام، والخطة، وظروف الوقت. [K04–K07]

المجموعة المستهدفة:

- "السكان العالميون"
- "مستخدمي الإنترنت",
- "مستخدمي الذكاء الاصطناعي",
- "العملاء"

لا يمكن ترك السكان تحت تسمية غامضة مثل 'المستخدمون المعنيون'. حسب الاقتضاء، يجب الإجابة على الأسئلة التالية: أي منتج ذكاء اصطناعي؟ أي واجهة المستخدم؟ أي الدول؟ الوصول اعتبارًا من أي تاريخ؟ ما هو شرط العمر؟ ما هي اللغة أو اللغات؟ أي جهاز وواجهة؟ أي خطة أو نوع حساب؟ ما هي الشروط الرسمية للاستخدام؟ هل يجب على المستخدمين الدفع للوصول إلى المنتج؟ هل ينطبق سؤال القياس على جميع المستخدمين أم على سوق أو ولاية قضائية معينة فقط؟ هل يتم قياس المستخدمين النشطين أم المستخدمين المحتملين؟ أي مجموعات مستخدمين خارج نطاق الدراسة؟ أي مجموعات تنتمي إلى السكان المستهدفين لكنها غائبة عن إطار العينة؟ قد يؤدي التغيير في أي من هذه الحقول إلى تغيير السكان المستهدفين. وعندما يحدث ذلك، لا يمكن دمج الدرجات القديمة والجديدة في نفس السلسلة دون توضيح.

3. السكان ليسوا رقمًا واحدًا

يعد إجمالي عدد السكان الرسمي أو المقدر للبلد مجرد الطبقة الابتدائية لـ GEO-1000. يمكن أن يتكون عدد السكان المراجع من الطبقات التالية المحددة.

3.1. السكان المقيمون

إجمالي عدد السكان البشريين الذين يعيشون في البلد أو الاختصاص المحدد. وهذا العدد:

- لا يعني أي ملاءمة من حيث العمر،
- الإنترنت،
- الوصول إلى المنتج،
- اللغة،
- الحساب،
- الجهاز

الاعتبارات.

3.2. السكان المؤهلون من حيث العمر

السكان الذين يستوفون شرط العمر للمنتج أو بروتوكول البحث المتعلق بالذكاء الاصطناعي. بالنسبة للمنتج:

- العمر الأدنى للاستخدام،
- متطلبات موافقة الوالدين،
- قاعدة العمر الخاصة بالدولة

قد توجد. قد يكون بروتوكول البحث أكثر تقييداً من شرط المنتج. على سبيل المثال، قد يتألف عينة GEO-1000 التأسيسية من مشاركين بالغين فقط لتقليل التعقيدات الأخلاقية والقانونية. في هذه الحالة، النتيجة ليست: "جميع السكان":

بل "السكان البالغين والمؤهلين للمنتج"

النتيجة.

3.3. السكان المتصلون

السكان الذين يمكنهم استخدام المنتج المعني بمستوى حيث:

- الإنترنت،
- الجهاز،
- جودة الاتصال،
- الوصول التقني

متاح. قد لا يكون الوصول إلى الإنترنت سمة ثنائية. الشخص:

- يمكنه استخدام المراسلة،
- ومع ذلك، قد لا يتمكن من تشغيل تطبيق الويب الكبير،
- قد يستخدم الإنترنت عبر الهاتف المحمول،
- ومع ذلك، قد لا يتمكن من الوصول إلى تجربة سطح المكتب.

قد لا يكون قادرًا على تحميل الردود أو لقطات الشاشة بسبب ضعف الاتصال. لا ينبغي أن تؤدي قيود الاتصال تلقائيًا إلى استبعاد المستخدم من الفئة المستهدفة. يجب أن تكون مرتبطة بظروف الاستخدام الفعلية للمنتج المقاس.

3.4. السكان القابلين للوصول من منظور المنتج

بالنسبة للمنتج الذكاء الاصطناعي ذي الصلة في التاريخ المحدد:

- تم إطلاقه رسميًا،
- يمكن استخدامه وفقًا للقانون وشروط المنتج،
- يمكن الوصول إلى واجهة الاستخدام اللازمة

إنها السكان. غياب منتج في بلد ما لا يعني أن الناس في ذلك البلد ممثلون تمثيلاً خاطئاً. في هذه الحالة: الرد على التمثيل هو لا شيء، لا يوجد تغطية للوصول. يجب الاحتفاظ بالنتيجتين بشكل منفصل.

3.5. السكان مناسبون من حيث الحساب والخطة

بعض المنتجات:

- قد تتطلب حسابًا،
- قد تتطلب التحقق عبر الهاتف،
- قد تتطلب طريقة دفع،
- قد يتم تقديمها فقط بموجب عقد شركة,

قد يفتح بعض الميزات فقط في خطة مدفوعة. هذه الشروط تؤثر على الجزء من السكان المستهدفين الذي يمكنه فعليًا الوصول إلى واجهة المستخدم ذات الصلة. لا يمكن تطبيق درجة الخطة المجانية تلقائيًا على مستخدمي الخطة المدفوعة.

3.6. السكان المناسبون من حيث اللغة

السكان الذين يمكنهم استخدام المُدخل والاستجابة للذكاء الاصطناعي ذات الصلة بمستوى كافٍ في لغة التدقيق.
مناسب اللغة فقط:

- لا يمكن تحديده فقط من خلال اللغة الرسمية للبلد.
- اللغة الأم للشخص

يجب التمييز بين:

- اللغة الأم أو لغة المنزل
- لغة الاستخدام اليومي
- لغة واجهة الذكاء الاصطناعي
- لغة المُدخل
- لغة الاستجابة
- اللغة للاستخدام المهني
- الكفاءة في القراءة

يمكن للشخص أن يكون مناسبًا في لغات متعددة. لا يمكن احتساب هذا الشخص كأكثر من شخص كامل واحد في حساب السكان العالمي.

3.7. السكان المؤهلون للواجهة

القياس:

- الويب،
- تطبيق الهاتف المحمول,
- تطبيق سطح المكتب,
- بيئة العمل المؤسسية,
- المساعد المدمج

يمكن استخدامها. قد يختلف عدد المستخدمين المتاحين لكل واجهة. قد لا يمتلك المستخدم الذي لديه وصول إلى الويب المنتج المحمول. قد لا يكون واجهة المستخدم الخاص بالشركة متاحًا للجمهور.

3.8. السكان المؤهلون للبحث

حتى إذا كان بإمكان الشخص استخدام المنتج، فإن للإدراج في البحث، بالإضافة إلى ذلك:

- الموافقة الطوعية،
- الخصوصية،
- معالجة البيانات،
- جمع لقطات الشاشة،
- متطلبات البحث الخاصة بالدولة،
- لا يعني أي ملاءمة من حيث العمر،
- الأمن

قد تحتاج الشروط إلى أن تتحقق. الأشخاص الذين لا يمكن إدراجهم في البحث ولكنهم يستخدمون المنتج الفعلي قد يكونون جزءًا من الفئة المستهدفة. في هذه الحالة، إطار العينة لا يمثل بشكل كافٍ الفئة المستهدفة. لا ينبغي إخفاء هذا النقص في التمثيل.

3.9. السكان المؤهلون تحليليًا

السكان الذين يمكن استخدام البيانات المجمعة لإنتاج نتائج لهم. المشارك:

- غيّر صياغة المُدخل،
- استخدم المنتج الخاطئ،
- اختر الإجابة،
- قدّم دليلاً غير مكتمل

من الممكن. حتى إذا كان الفرد ينتمي إلى السكان المستهدفين، قد لا يتم إدخال ملاحظته في مجموعة البيانات التحليلية. الملاحظات الصالحة والسكان المستهدفون ليسوا الشيء نفسه.

4. تعريف سكان المستخدمين المناسبين

لمنتج الذكاء الاصطناعي a في الوقت t، يمكن تعريف الكون المؤهل للمستخدمين على النحو التالي:

```
U_{a,t}= {u ∈ P_t: AgeEligible(u,a,t)=1∧ Connected(u,a,t)=1∧ ProductAvailable(u,a,t)=1∧
LawfulAccess(u,a,t)=1∧ InterfaceEligible(u,a,t)=1}
```

هنا:

- Pt: السكان البشريون العامون في التاريخ المحدد
- u: الفرد البشري
- a: منتج الذكاء الاصطناعي و واجهة المستخدم
- t: وقت القياس

من حيث اللغة والتعليمات، يمكن لاحقًا تضييق الكون المستهدف على النحو التالي:

```
U_{a,c,l,s,p,t}= {u∈ U_{a,t}: Country(u)=c ∧ LanguageEligible(u,l)=1∧ SurfaceEligible(u,s)=1∧
PromptRelevant(u,p)=1}
```

هنا:

- c: الدولة أو الولاية القضائية
- l: اللغة والمحلية
- s: واجهة المستخدم
- p: المُدخل أو عائلة الأسئلة

الغرض من هذا الترميز ليس تسجيل جميع الأفراد واحدًا تلو الآخر. بل هو لجعل الحدود المنطقية للفئة المستهدفة واضحة.

5. ضرب النسب الحدية بشكل أعمى ممنوع

يمكن أحيانًا حساب الفئة المؤهلة تقريبًا باستخدام صيغة مثل التالية:

```
N× p_age× p_internet× p_product× p_language
```

لا يمكن تبرير هذا الحساب إلا تحت استقلالية الشروط أو افتراض وجود توزيع مشترك مناسب. في الحياة الواقعية، قد لا تكون الخصائص مستقلة. على سبيل المثال:

- استخدام الإنترنت قد يختلف حسب العمر،
- الوصول إلى خطة مدفوعة قد يكون مرتبطًا بالدخل،
- استخدام لغة معينة قد يكون مرتبطًا بالمنطقة داخل البلد والتعليم،

الوصول إلى الجهاز وجودة الاتصال قد تتغير معًا. ضرب المعدلات الهامشية يمكن أن يتنبأ بسكان الهدف الذين هم:

- أكبر من اللازم،
- أصغر من اللازم

يجب أن يكون ترتيب الأفضلية كالتالي:

- بيانات ميكروية مشتركة مناسبة أو جدول تقاطعي
- توقع سكاني متعدد الأبعاد وموثوق
- نموذج معاير
- نهج هامشي مع شرح المبررات وتحليل الحساسية

إذا تم استخدام الضرب البسيط:

- افتراض،
- مصدر البيانات،
- عدم اليقين،
- سيناريو بديل

يجب الإفصاح عنها.

6. سؤال التدقيق يغير السكان المستهدفين

ليس كل مُدخل له نفس universo القياس لجميع سكان العالم.

6.1. أسئلة الهوية العالمية

مثال: "ما نوع الشركة Apple.com؟" يمكن طرح هذا السؤال على نطاق واسع بين المستخدمين العالميين المناسبين للمنتج. لأن هوية الشركة لا تقتصر على مستخدمي سوق محدد.

6.2. أسئلة الخدمة المحلية

مثال: "هل تقدم هذه الشركة خدمات في تركيا؟" الفئة المستهدفة الرئيسية:

- المستخدمون المؤهلون في تركيا،
- المستخدمون الخارجيون الذين يبحثون عن خدمة تركيا

هذا ممكن. ليس من الضروري إعطاء الوزن نفسه لجميع السكان في العالم.

6.3. أسئلة الأسعار والشراء

السعر:

- البلد،
- الضرائب،
- عملة
- الخطة،
- نوع العميل

قد يختلف. يجب تحديد الفئة المستهدفة وفقًا للسوق التجاري المعني ونوع المستخدم.

6.4. القضايا القانونية والترخيصية

يعتمد الترخيص والسلطة القانونية على الولاية القضائية. الفئة المستهدفة:

- يمكنها الحصول على الخدمة ذات الصلة في ذلك البلد،

- وتخضع لذلك النظام القانوني.

قد تكون الفئة المستهدفة محدودة وفقًا لذلك.

6.5. القضايا الاستشارية

الفئة المستهدفة للسؤال "هل هذه الشركة مناسبة لي؟" ليست جميع الأشخاص. يمكن تحديدها وفقًا لـ:

- احتياجات المستخدم
- البلد
- الميزانية
- نوع العميل
- نطاق الخدمة
- شروط الأهلية والاستبعاد

قد تكون الفئة الاستشارية أضيق من الفئة المحددة بالهوية.

6.6. الأسئلة عالية المخاطر

في مجالات الصحة أو القانون أو المالية أو الأمن:

- الحكم,
- ملف المستخدم,
- حالة المخاطر,
- السلطة المهنية

تحد بشكل أقوى من الفئة المستهدفة.

7. يجب ربط إطار السكان بعائلة صياغات المُدخل

في التدقيق، قد تكون هناك أطر سكانية متعددة بدلاً من وجود مجموعة مستهدفة واحدة.

7.1. السكان الأساسيون للهوية

السكان العالميون أو المؤهلون بشكل واسع لقياس هوية الكيان الأساسية.

7.2. السكان ذوو الصلة بالخدمة

هم السكان الذين يمكن للكيان تقديم الخدمة لهم فعليًا أو يمكن اعتباره مؤهلاً للحصول على الخدمة من قبل المستخدم.

7.3. السكان المعنيون بالمعاملات السوقية

السكان الذين يمكن تطبيق السعر أو عملية الشراء أو الإرجاع أو الشرط التجاري المحدد عليهم.

7.4. السكان القضائيون

إنها فئة المستخدمين التي يمكن تطبيق ترخيص محدد أو ادعاء قانوني أو تنظيمية عليها.

7.5. الفئة المؤهلة للتوصية

إنها فئة المستخدمين التي قد تكون مناسبة ضمن النطاق الحقيقي للكيان.

7.6. فئة المستخدمين المتأثرة عالية المخاطر

إنها فئة يمكن أن يتسبب التمثيل غير الصحيح في إلحاق ضرر كبير بها. يجب عدم الخلط بين هذه الأطر بعضها البعض. نتيجة الهوية العالمية لا تعني "موصى به للعالم بأسره."

8. نوع منتج الذكاء الاصطناعي يحدد الفئة

ليست كل منتجات الذكاء الاصطناعي منتجات دردشة متاحة للجمهور.

8.1. منتج الذكاء الاصطناعي للمستهلك العام

المجموعة المستهدفة:

- قد يكون هناك أشخاص يمكنهم الوصول إلى المنتج،
- وفقا لشروط الاستخدام،
- واستخدامه باللغة ذات الصلة.

يمكن تطبيق لوحة المستخدمين الحقيقية GEO-1000 مباشرة على هذا النوع من المنتجات.

8.2. منتج الذكاء الاصطناعي للشركات

المنتج مخصص فقط لـ:

- عقد صاحب العمل،
- رخصة الشركة،
- يمكن استخدامه بحساب مستخدم داخلي

الفئة المستهدفة ليست عامة البالغين:

```
إنه المقعد أو كون المستخدم مع الوصول المؤسسي إلى المنتج.
```

لا يمكن مقارنة النتائج المؤسسية بالمنتج الاستهلاكي العام حسب ترجيح السكان.

8.3. منتج تعليمي

المجموعة المستهدفة:

- طالب,
- معلم,
- مؤسسة,
- فئة عمرية,
- ترخيص تعليمي

قد يكون محدودًا من حيث. إذا كان من المقرر تضمين الأطفال أو القصر، فهناك حاجة إلى بروتوكول أخلاقي منفصل.

8.4. منتج للمطور أو API

قد لا يتم عرض إخراج API مباشرة من خلال واجهة المستخدم النهائية. الوحدة المستهدفة هي:

- مطور،
- تطبيق،
- طلب API،

- جلسة اختبار مراقبة

قد تكون. لا يمكن تطبيق لوحة السكان البشرية GEO-1000 مباشرة وبغير تغيير على API. بشكل منفصل لـ API:

تدقيق نظام مراقب

قد يكون مطلوبًا.

8.5. مساعد AI المدمج

الذكاء الاصطناعي:

- نظام التشغيل،
- واجهة البحث،
- منصة التسوق،
- برنامج الأعمال،
- الجهاز

قد يكون مضمنًا. السكان المستهدفون هم السكان الفعليون المستخدمون للمنتج المضمن. ليسوا إجمالي السكان المستخدمين لمزود النموذج الأساسي.

8.6. خبير أو ذكاء اصطناعي قطاعي

المنتج مخصص فقط لـ:

- محامون،
- أطباء،
- محترفو المالية،
- مهندسون

قد يتم تصميمه لهم. يجب أن يكون السكان المستهدفون محدودين بالمهنة وظروف الوصول.

9. حالات الوصول للمنتج

يجب تسجيل حالة الوصول لكل دولة، منتج، واجهة، وتاريخ.

PA-0— غير معروف

لا يمكن تحديد حالة الوصول بشكل موثوق.

PA-1— غير متاح

المنتج غير متاح في الدولة أو مجموعة المستخدمين ذات الصلة.

PA-2— مقيد

المنتج محدود على بعض:

- الحساب،
- الدعوات،
- مؤسسة,
- لا يعني أي ملاءمة من حيث العمر،

- الخطة،
- الجهاز

الشروط.

PA-3— متاح للوصول ولكن غير مدعوم رسميًا

يمكن ملاحظة الوصول الفعلي إلى المنتج. ومع ذلك، فإن حالة الدعم أو الاستخدام الرسمي غير واضحة أو محدودة. لا ينبغي إضافة هؤلاء المستخدمين تلقائيًا إلى اللوحة الرسمية الرئيسية للسكان. يمكن إنشاء لوحة تجريبية منفصلة.

PA-4— متاح رسميًا

المنتج متاح رسميًا وتُحدد شروط الاستخدام الرئيسية.

PA-5— متاح رسميًا مع واجهة تدقيق كامل

المنتج ليس غير متاح فقط؛ بل هو خاضع للرقابة:

- اللغة،
- الخطة،
- واجهة
- تحت شروط الجلسة

مناسب لتوليد الملاحظات المطلوبة بواسطة البروتوكول. توفر المنتج في بلد ما لا يعني أن جميع الخطط واللغات متوافقة مع السيطرة.

10. تجاوز القيود ليس قياسًا حقيقيًا للسكان

لا يمكن طلب من المشاركين تنفيذ السلوكيات التالية:

- إعلان عن البلد الخطأ
- إنشاء حساب مزيف
- تغيير معلومات العمر
- تجاوز قيود الوصول الرسمية
- استخدام حساب شخص آخر
- تجاوز أمان المنتج أو التحقق منه
- انتهاك شروط الاستخدام

يمكن أيضاً فحص هذه السلوكيات كتجربة تقنية. ومع ذلك:

تجربة المستخدم الطبيعي والمشروع

لا يمكن الإبلاغ عنها على هذا النحو. إذا لم يكن هناك وصول في بلد ما، يجب أن تكون الحالة الصحيحة: غير متاح / أو / خارج الوصول الأصلي. الملاحظة الفردية التي تتجاوز القيد لا تولد درجة سكان البلد.

11. التمييز بين المستخدم المحتمل والمستخدم النشط

الأشخاص الذين يمكنهم استخدام المنتج ليسوا نفس المجتمع الذي يستخدم المنتج فعليًا.

11.1. السكان ذوو الوصول المحتمل

هؤلاء هم الأشخاص الذين يمكنهم استخدام المنتج تقنيًا وقانونيًا. هذا السكان يجيب على السؤال: 'ماذا سيرى الشخص المناسب إذا استخدم المنتج؟' السكان ذو قيمة من أجل العدالة ومقارنة المنتجات.

11.2. عدد المستخدمين النشطين

هؤلاء هم الأشخاص الذين استخدموا المنتج فعليًا خلال فترة محددة. يقترب هذا العدد من السؤال: 'ماذا يرى الناس ضمن حجم الاستخدام الفعلي؟' ويتطلب بيانات موثوقة عن الاستخدام النشط.

11.3. عدد المستخدمين المعرضين للاستعلام

هؤلاء هم الأشخاص الذين يستخدمون المنتج ولديهم إمكانية طرح أسئلة حول العلامة التجارية أو الموضوع الذي يتم مراقبته. غالبًا ما لا يمكن معرفة هذا العدد مباشرة. قد تكون بيانات سلوك البحث أو الاستعلام مطلوبة.

11.4. عدد الأشخاص الذين يرون الاستجابة بالفعل

هم المستخدمون الذين يتعرضون فعليًا لإخراج تمثيل محدد. قد لا يكون من الممكن قياس ذلك بالكامل بدون بيانات مزود المنتج الفعلية خارج لوحة المستخدمين.

12. وزن التعرض ليس نفس وزن الأهلية

يمكن توليد نتيجتين عالميتين منفصلتين.

12.1. نتيجة الوزن السكاني للأهلية

يتم وزن كل شخص وفقًا لحصته ضمن السكان المؤهلين المحددين. هذه النتيجة هي:

- عالمياً عادلة،
- ربما قابلة للوصول،
- تمثل السكان

من حيث

12.2. النتيجة المرجحة بالتعرض الفعلي

يتم ترجيح منتجات الذكاء الاصطناعي أو مجموعات المستخدمين وفقًا لأحجام الاستخدام الفعلية. هذه النتيجة:

- ذات قيمة من حيث
- التأثير التجاري الحالي،
- التعرض الفعلي للمستخدم،

المخاطر التشغيلية.

ومع ذلك، قد لا تتوفر بيانات استخدام موثوقة. قد تعتمد التقديرات على تصريحات المزود أو بيانات السوق أو عينات غير مكتملة. لذلك يجب توضيح نوع الترجيح بشكل صريح.

12.3. يُحظر استبدال واحد بآخر

درجة السكان المؤهلين: لا تعني "معظم المستخدمين الفعليين رأوا هذا الرد." درجة التعرض: لا تعني "يتم ضمان تمثيل عادل للسكان في العالم."

13. يجب فصل نطاق الوصول ودقة التمثيل

قد يكون منتج الذكاء الاصطناعي متاحًا فقط في دول محدودة. في هذه الدول، يمكنه إنتاج تمثيلات دقيقة جدًا. قد يكون منتج آخر متاحًا على نطاق أوسع لكنه قد ينتج دقة أقل. إذا نظرنا فقط إلى دقة التمثيل، قد يبدو أن المنتج محدود الوصول متفوقًا. إذا نظرنا فقط إلى تغطية الوصول، قد يتم التغاضي عن التمثيل غير الصحيح. لذلك، هناك حاجة إلى نتيجتين منفصلتين.

13.1. تغطية الوصول

يشير إلى مدى قدرة شريحة السكان العالمية المحددة على استخدام المنتج الذكاء الاصطناعي المعني بشكل مناسب. تمثيل المرشح:

```
AC_a= N(U_{a,t}) / N(G_t)
```

هنا:

- $U_{a,t}$: عدد المستخدمين المؤهلين للمنتج
- G_t: السكان المرجعيون العالميون المحددون

13.2. دقة التمثيل الشرطي

هي احتمالية رؤية تمثيل دقيق بين المستخدمين الذين يمكنهم الوصول إلى المنتج. تدوين المرشح:

```
RA_a= Pr(Y=1| u ∈ U_{a,t})
```

13.3. الوصول الفعال إلى التمثيل الدقيق

مفهوم المرشح

يمكن تطوير مقياس تشغيلي يجمع بين تغطية الوصول والدقة الشرطية:

```
ECR_a= AC_a× RA_a
```

يقارب هذا المقياس السؤال: "ما النسبة من السكان المرجعيين العالميين المحددين الذين يمكنهم الوصول إلى هذا المنتج وتلبية شرط رؤية تمثيل دقيق؟" هذه ليست بعد الصيغة النهائية NOMOS. لأن:

- الوصول مختلف عن الاستخدام الفعلي،
- الوزن وهياكل الاعتماد معقدة،
- استخدام المنتج طوعي،

التمثيل الصحيح يعتمد على صياغة المُدخل. ومع ذلك، فإنه يحافظ على التمييز الأساسي:

> عدم القدرة على الوصول ليس هو نفسه الحصول على إجابة خاطئة.

14. نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة

عند مقارنة منتجات الذكاء الاصطناعي، هناك حاجة إلى عالمي سكان منفصلين.

14.1. سكان نطاق الوصول الطبيعي

لكل منتج ذكاء اصطناعي جمهوره المؤهل ضمن نطاق الوصول الرسمي والفعلي له. هذه النتيجة تجيب على السؤال التالي:

> إلى أي مدى يمثل المنتج بدقة السكان الذين يمكنه الوصول إليهم فعليًا؟

قد تختلف مجموعات السكان المستهدفة بين المنتجات.

14.2. سكان قاعدة المقارنة المشتركة

في شكل مشترك عبر جميع منتجات الذكاء الاصطناعي المقارنة:

- متاح رسميًا،
- يمكن استخدامه تحت نفس شرط اللغة أو واجهة المستخدم

هو السكان. هذا الناتج يجيب على السؤال التالي:

> أي منتج ينتج تمثيلًا أكثر دقة في نفس كون المستخدمين؟

14.3. يجب نشر نتيجتين معًا

نطاق الوصول الطبيعي:

- يُظهر الوصول الفعلي للمنتج،
- التوزيع الجغرافي الطبيعي

قاعدة المقارنة المشتركة: يوفر مقارنة أكثر عدلاً للمنتجات. قد يكون المنتج مرتفعًا في درجة نطاق الوصول الطبيعي، لكنه أقل في درجة قاعدة المقارنة المشتركة. نتيجتان تجيبان على أسئلة مختلفة.

15. السكان المرتبطون بالسوق

قد يختلف مؤشر الهوية العالمية لكيان ما عن مؤشّر سوقه التجاري المستهدف. مثال: إذا كانت شركة محاماة محلية تقدم خدماتها فقط في تركيا: يمكن قياس سؤال الهوية العالمية في إجمالي السكان المؤهلين حول العالم، ولكن يجب قياس التوصيات ومدى ملاءمة الخدمات فقط للمستخدمين في تركيا أو ضمن نطاق الخدمات الفعلي. يمكن للمؤسسة اختيار النتيجة لصالحها:

- السوق ذا الدرجة العالية فقط،
- لغته/لغتها فقط،
- عملائها الحاليون فقط

لا يمكن الإعلان عن السكان المستهدفين. السوق المستهدفة:

- قبل بدء التدقيق،
- وفق سجلات الخدمة،
- ضمن نطاق العقد والعمليات

يجب أن يتم تثبيته وفقًا.

16. فصل الشخص والحساب والجلسة

وزن السكان ينتمي إلى الناس. وليس إلى حسابات الذكاء الاصطناعي.

16.1. شخص واحد، حسابات متعددة

نفس الشخص:

- مجاني،
- مدفوع،
- شركات

يمكن أن يكون لدى المستخدمين عدة حسابات. يمكن لهذه الحسابات أن تنتج ملاحظات منفصلة في مقارنات واجهة المنتج. لا يمكن احتسابها كأفراد منفصلين في سكان البشر العالمي.

16.2. حساب واحد، عدة أشخاص

قد يتم استخدام حساب عائلي أو فريق أو حساب شركة من قبل عدة أشخاص. لا يساوي ذلك عدد مستخدمي الحساب.

16.3. شخص واحد، جلسات متعددة

يمكن للشخص أن ينتج ملاحظات منفصلة في جلسات مختلفة. يمكن أن تقيس هذه الملاحظات الوقت أو استقرار الجلسة. إنها لا تزيد من حجم السكان المستقل.

16.4. شخص واحد، منتجات ذكاء اصطناعي متعددة

إنه ذو قيمة لمقارنة المنتجات المزدوجة. ومع ذلك، الملاحظات تعتمد على بعضها بسبب الخصائص المشتركة على مستوى المستخدم.

16.5. وحدة الوزن الرئيسية

في درجة GEO القائمة على السكان، يجب أن تكون وحدة الوزن الرئيسية:

بشرية

يجب ربط الحسابات والجلسة والملاحظات الخاصة بالمنتج تحت وحدة الإنسان.

17. تعداد اللغة

استخدام لغة الدولة الرسمية فقط ليس كافيًا لتحقيق العدالة اللغوية.

17.1. لغة المُدخل

اللغة التي يطرح بها المشارك سؤال التدقيق.

17.2. لغة الواجهة

اللغة المستخدمة في قائمة وواجهة مستخدم منتج الذكاء الاصطناعي. قد لا تكون نفسها لغة المُدخل.

17.3. لغة الاستجابة

اللغة التي يجيب بها الذكاء الاصطناعي فعليًا. قد يستجيب النظام بلغة مختلفة عن لغة المُدخل. هذا أيضًا نتيجة قياس.

17.4. لغة كفاءة المستخدم

اللغة التي يمكن للمشارك استخدامها بشكل مفيد في المُدخل والاستجابة. لا يلزم أن تكون اللغة الأم.

17.5. الإقليم

السياق الإقليمي إلى جانب اللغة. يمكن أن تكون نفس اللغة مختلفة:

- البلد،
- السعر،
- القانون،
- المصطلحات

قد تحمل الشرط.

18. العد المزدوج للمستخدمين متعددين اللغات

شخص:

- التركية،
- قد يكون قادرًا على استخدام اللغة الإنجليزية.
- الألمانية

قد يخضع هذا الشخص لثلاثة اختبارات لغوية منفصلة. ومع ذلك، احتسابه كثلاثة أشخاص في وزن سكان العالم يُشَوِّه تقدير السكان. يمكن استخدام ثلاث طرق.

18.1. تخصيص لغة أم واحدة

يتم تعيين الشخص لخانة لغة واحدة في التحليل الرئيسي. تُستخدم اللغات الأخرى كاختبارات إضافية أو مزدوجة.

18.2. تقسيم وزن الشخص

يمكن تقسيم وزن الشخص السكاني بين اللغات المناسبة بطريقة محددة مسبقًا. لا يتجاوز الوزن الكلي شخصًا واحدًا.

18.3. لوحة متعددة اللغات متطابقة منفصلة

يُستخدم المستخدمون متعددو اللغات للمقارنات بين اللغات للشخص نفسه. يتم تحليل هذه العينة بشكل منفصل عن عينة السكان الرئيسية. بغض النظر عن الطريقة المستخدمة:

> لا يمكن تكرار الوزن الكلي للسكان للشخص نفسه دون تفسير.

19. كيف يجب التحقق من كفاءة اللغة؟

تصريح أحد المشاركين: "أنا أعرف هذه اللغة." يمكن أن يكون سجلًا أوليًا. ومع ذلك، في التدقيق عالي الأمان، يمكن استخدام إحدى الطرق التالية:

- إعلان اللغة الأم أو الاستخدام اليومي
- اختبار قصير للقراءة والفهم
- سجل اللغة لمقدم العينة
- التحقق من الحكم المحلي
- سجل الاستخدام التعليمي أو المهني
- نجاح مهمة تجريبية

لوحات اختبارات اللغة الصعبة للغاية يمكن الوصول إليها للمستخدمين المدربين وذوي الدخل العالي. الهدف ليس إجراء اختبار لغة أكاديمي. الهدف هو التحقق من الكفاءة لفهم صياغة المُدخل والاستجابة كمستخدم حقيقي.

20. ماذا يحدث إذا لم يتم دعم لغة معينة؟

قد يقوم منتج الذكاء الاصطناعي بما يلي:

- عدم دعم لغة معينة رسميًا،
- دعمها جزئيًا،
- عدم الادعاء بدعمها لكنه يمكنه توليد استجابات

قد يكون كذلك. يجب تسجيل حالات اللغة بشكل منفصل:

مدعوم رسميًا

دعم محدود

ملاحظ ولكن غير مدعوم

غير مدعوم

مجهول

تم استلام الرد بلغة غير مدعومة رسميًا:

- اكتشاف،
- وصول تجريبي،
- صلابة اللغة

قد يتم قياسه. لا ينبغي إضافته تلقائيًا إلى نتيجة عدد المستخدمين المدعومين الرئيسيين. يجب أيضًا الإبلاغ عن حد الوصول إلى اللغة.

21. وزن سكان الدولة وتغطية مراقب الدولة

تخدم اللوحة الرئيسية الموزونة حسب السكان ولوحة رؤية الدولة أغراضًا منفصلة.

21.1. لوحة السكان

يتم توزيع المشاركين وفقًا لنصيبهم في السكان المستهدفين المناسبين للمنتج. تتلقى التجمعات السكانية الأكبر المؤهلة وزنًا أعلى. تنتج النتيجة الرئيسية للسكان العالميين.

21.2. لوحة مراقب الدولة

تسعى لوحة مراقب الدولة للحصول على ملاحظة واحدة على الأقل من الدول الصغيرة التي تحصل على تمثيل قليل أو معدوم في عينة السكان الرئيسية. تشير ملاحظة مراقب الدولة الفردية إلى:

- أن الوصول متاح،
- وأن نتيجة معينة وفردية قد حدثت.
- يُظهر الحاجة إلى أبحاث جديدة

قد يشير إلى ذلك. لا ينتج درجة للدولة بمفرده.

21.3. كفاية للحصول على درجة الدولة

لنُشر معدلات تعتمد على الدولة:

- الحد الأدنى للعينة،
- معدل الملاحظة الصالح،
- عدم اليقين،
- تمثيلية العينة

يجب تلبية الشروط. مع وجود مشارك واحد أو عدة مشاركين: 'درجة الدولة NOMOS' لا يمكن الإعلان عنها.

22. عينة إنصاف اللغة

قد يجعل وزن السكان اللغات ذات السكان الأقل أو الموارد القليلة غير مرئية في الدرجة الرئيسية. لذلك، يمكن إنشاء عينة منفصلة لإنصاف اللغة. الغرض:

- لاختبار اللغات التي تمثل أقلية،
- لاكتشاف التدهور الشديد القائم على اللغة،
- لتحديد مشكلات ترجمة المُدخلات،
- لجعل سلوك اللغة منخفضة الموارد مرئياً

يجب أن يكون. يمكن لعينة عدالة اللغة تقديم عينات إضافية إلى عينة السكان الرئيسية. ومع ذلك، يجب إعادة ملاحظات اللغة التي تم أخذ عينات زائدة لها إلى أوزان السكان الفعلية في النتيجة العالمية. وإلا فإن عينة العدالة ستغير تقدير السكان العالمي.

23. عينة الوصول

المستخدمون الذين لديهم احتياجات بصرية أو سمعية أو حركية أو معرفية أو غيرها تتعلق بإمكانية الوصول هم جزء حقيقي من السكان المستهدفين. إذا لم يكن مهمة الالتقاط أو البحث NOMOS مناسبة لمشاركة هؤلاء الأفراد: لا يمكن إزالة المستخدمين من السكان المستهدفين. النتيجة الصحيحة:

```
إطار العينة غير مكتمل من حيث إمكانية الوصول.
```

يمكن للوحة وصول منفصلة تقييم ما يلي:

- توافق مع قارئ الشاشة
- استخدام لوحة المفاتيح
- الإدخال الصوتي
- إعدادات ضعف البصر
- الحمل المعرفي
- إمكانية الوصول إلى الهاتف المحمول
- قابلية استخدام أداة الالتقاط

ينبغي أن يرى لوحة الوصول المستخدمين كجزء من تغطية السكان، وليس كـ "استثناء خاص".

24. السكان المستهدفون وإطار التحليل العيني

لا ينبغي الخلط بين هذين المفهومين.

24.1. السكان المستهدفون

المجموعة الكاملة من الأشخاص التي يرغب المرء في الوصول إلى استنتاج بشأنها.

24.2. الإطار العيني

السكان الذين يمكن فعليًا اختيار المشاركين منهم:

- لوحة,
- قائمة,
- قاعدة بيانات

- منظمة مشتركة
- مجموعة المستخدمين

هو العنقود. قد لا يغطي إطار العينة كامل السكان المستهدفين.

24.3. السكان القابلون للاتصال

هؤلاء هم الأشخاص الذين يمكن فعليًا توصيل دعوة البحث لهم.

24.4. السكان الراغبون في المشاركة

أولئك الذين يشاركون طوعًا في البحث.

24.5. السكان الذين يكملون المهمة

أولئك الذين يكملون مهمة التدقيق.

24.6. السكان التحليليين الصالحين

هم الأشخاص الذين ينتجون ملاحظات مناسبة للبروتوكول وقابلة للاستخدام في التحليل. يجب الإبلاغ عن أعداد هذه الطبقات بشكل منفصل.

25. خطأ التغطية

قد يستبعد إطار العينات بعض أجزاء السكان المستهدفين. يُطلق على ذلك خطأ التغطية. أمثلة:

- عينات مهنية على الإنترنت فقط
- مستخدمي أجهزة الكمبيوتر المكتبية فقط
- الأشخاص الذين لديهم بطاقات ائتمان فقط
- أولئك الذين يستخدمون واجهة باللغة الإنجليزية فقط
- سكان المدن الكبيرة فقط
- المستخدمون المهتمون بالتكنولوجيا بشكل كبير فقط
- استبعاد الأشخاص ذوي احتياجات الوصول

خطأ التغطية:

- يجب شرحه،
- يجب قياسه إذا أمكن

يجب تقليله باستخدام الترجيح أو لوحة إضافية. الترجيح لا يمكنه خلق مجموعة غير موجودة في إطار العينة بشكل سحري.

26. التغطية الزائدة

قد يتم العثور على أشخاص لا ينتمون إلى السكان المستهدفين في إطار العينة. أمثلة:

- مشترك في بلد حيث المنتج غير متاح
- شخص لا يستوفي شرط العمر المطلوب
- خطة غير صحيحة أو واجهة المستخدم
- مشترك لا يستطيع استخدام لغة التدقيق

- شخص تم العثور عليه مرة أخرى بتسجيلات متعددة
- مستخدم يعمل نيابة عن دولة أخرى

يجب ألا يتم تضمين سجلات التغطية الزائدة في الدرجة الرئيسية. ومع ذلك، يجب الإبلاغ عن سبب الاستبعاد والمعدل.

27. عدم الاستجابة والاختيار الذاتي

قد يختلف الأشخاص الذين يوافقون على المشاركة في البحث بشكل منهجي عن السكان المستهدفين. المتطوعون:

- قد يستخدمون منتجات الذكاء الاصطناعي بشكل أكثر تكرارًا،
- قد يكونون أكثر اهتمامًا بالتكنولوجيا،
- قد يكون لديهم مستوى تعليم أعلى،
- قد يكونون بالفعل على دراية بالعلامة التجارية,

قد يتصرف بشكل مختلف بسبب المكافأة أو التعويض. لذلك: يجب استخدام التعبير "1,000 أشخاص عشوائيون في العالم" فقط في تصميمات الاختيار الاحتمالية والموثقة حقًا. إذا تم استخدام متطوع أو لوحة عبر الإنترنت، فإن التعبير الصحيح هو:

لوحة مستخدمين اختيارية متعددة الطبقات ومرجحة

قد يكون ذلك ممكنًا. لا يمكن تخزين نوع اللوحة.

28. يتغير السكان المناسبون مع مرور الوقت

كون المستخدمين لمنتج الذكاء الاصطناعي ليس ثابتًا. يمكن أن تتغير الأمور التالية:

- الدول المدعومة
- الحد الأدنى للعمر
- الوصول المجاني
- هيكل خطط الدفع
- التوفر على الهاتف المحمول أو الويب
- دعم اللغة
- شروط الحساب
- تغطية منتجات الشركات
- الوضع القانوني والتنظيمي

لهذا السبب، يجب أن يحتوي كل سجل سكاني على:

- تاريخ الملاحظة،
- فترة الصلاحية،
- نسخة المصدر،
- آخر تاريخ تحقق

يجب أن يتحمل.

28.1. التغيير خلال موجة القياس

إذا تغير الوصول إلى المنتج جوهريًا خلال الموجة:

- يمكن إيقاف الموجة،
- يمكن إنشاء فترتين منفصلتين قبل وبعد التغيير.

يمكن تحليل الدول المتأثرة بشكل منفصل. لا يمكن دمج المستخدمين قبل وبعد التغيير كسكان موحدين.

29. ملف مصدر السكان

يجب على كل تدقيق تخزين بيانات السكان المستهدفين التي يستخدمها في ملف بإصدار محدد. وعند الاقتضاء، يجب تضمين الحقول التالية:

- البلد
- السكان المرجعيون
- تاريخ السكان
- أهلية العمر
- الوصول إلى الإنترنت والأجهزة
- حالة الوصول إلى المنتج
- وصول الواجهة
- متطلبات الخطة والحساب
- مجموعات اللغة
- أهلية البحث
- طريقة التوزيع الشائعة
- عدم اليقين
- المصادر
- تاريخ آخر تحقق
- الشخص المسؤول
- نسخة السكان

لا يمكن إعادة إنتاج حصص الدول بعد النتائج بدون هذا الملف.

30. عرض المنتج الاصطناعي والدولة

عرض المنهجية الاصطناعية / الدول التالية، المنتج بالذكاء الاصطناعي، أعداد السكان، ومعدلات الوصول كلها خيالية تمامًا. لا تمثل أي دولة أو منتج حقيقي. دعونا نعتبر منتجًا استهلاكيًا عامًا اصطناعيًا باسم AI Orion. هناك أربع دول اصطناعية.

البلد	إجمالي السكان	السكان المناسبون المقدّرون بعد عوامل العمر، الإنترنت، اللغة، وواجهة الاستخدام	وصول المنتج	عدد سكان نطاق الوصول الطبيعي
البلد أ	1,400,000,000	780,000,000	غير متاح	0
البلد ب	1,450,000,000	520,000,000	الوصول الرسمي	520,000,000
البلد ج	340,000,000	250,000,000	الوصول الرسمي	250,000,000

البلد	إجمالي السكان	السكان المناسبون المقدّرون بعد عوامل العمر، الإنترنت، اللغة، وواجهة الاستخدام	وصول المنتج	عدد سكان نطاق الوصول الطبيعي
البلد د	35,000	25,000	الوصول الرسمي	25,000
الإجمالي	3,190,035,000	1,550,025,000		770,025,000

30.1. إذا تم استخدام إجمالي عدد السكان

سيكون للبلد أ حصة كبيرة في العينة الرئيسية بسبب عدد سكانه الكبير جدًا. ومع ذلك، لا يمكن استخدام المنتج في ذلك البلد. لم يتمكن المشاركون من إنتاج تجربة مستخدم طبيعية.

30.2. عند استخدام عدد السكان نطاق الوصول الطبيعي

الحصص النظرية للسكان في لوحة المنتج الرئيسية تقريبًا:

- البلد ب: 67.53%
- البلد ج: 32.47%
- البلد د: 0.003%
- البلد أ: خارج نطاق الوصول الطبيعي

الحصة النظرية للبلد د في عينة رئيسية مكونة من 1,000 شخصًا أقل بكثير من شخص واحد. لذلك، يجب ألا تعطي النتيجة الرئيسية الموزونة بالسكان وزنًا كبيرًا للبلد د. ومع ذلك، إذا لم يكن الهدف أن يكون البلد غير مرئي تمامًا:

لوحة مراقبي البلدان

يمكن الحصول على ملاحظة استكشافية واحدة على الأقل. هذه الملاحظة:

- ليست نتيجة البلد د،
- لا تغيّر الوزن السكاني الرئيسي.

يُظهر أن سلوك المنتج يمكن مراقبته في البلد.

30.3. كيف يتم الإبلاغ عن البلد أ؟

بالنسبة للبلد أ:

يُستخدم وضع OUTSIDE_NATIVE_REACH

يجب ألا يخفي استبعاد البلد أ حدود الوصول إلى المنتج. كما تُظهر بطاقة النتائج العامة أيضًا:

- تغطية الوصول إلى المنتج
- الحصة غير المتاحة في السكان المرجعيين العالميين
- درجة نطاق الوصول الطبيعي
- درجة مقارنة على أساس مشترك

31. تمثيل السكان التجريبي الاصطناعي

مثال اصطناعي

لنأخذ بعين الاعتبار ثلاثة صياغات مُدخل لنفس منتج أوريون AI.

صياغة المُدخل أ

"ما نوع الشركة أستيرون؟" الفئة المستهدفة: يمكن أن تكون الفئة العريضة المناسبة لمستخدمي أوريون AI.

صياغة المُدخل ب

"هل تقدم أستيرون خدمات السفر للشركات في تركيا؟" الفئة المستهدفة:

- المستخدمون المؤهلون في تركيا،
- المستخدمون الخارجيون المحددون الذين يبحثون عن خدمة تركيا

قد يكون.

صياغة المُدخل ج

"هل أستيرون مناسب للاستشارات القانونية للهجرة بالنسبة لي؟" إذا لم تقدم أستيرون مثل هذه الخدمة: قد لا يكون هناك أي فئة مؤهلة للحصول على نصيحة، ويجب أن يكون الجواب الصحيح باستبعاد جميع المستخدمين المناسبين. دمج هذه صياغة المُدخل في درجة الهوية العامة قد يكون غير صحيح. هذا المثال يوضح:

> حتى بالنسبة لكائن واحد ونفس منتج الذكاء الاصطناعي، قد يختلف الجمهور المستهدف بناءً على السؤال الذي يقيسه المستخدم.

32. هندسة لوحة رباعية

هندسة مرشحة — هندسة مرشحة

يوصى بأربع لوحات تكميلية للتنفيذ العالمي والعادل لـ GEO-1000.

32.1. لوحة السكان

اللوحة الرئيسية مُوزونة وفقًا لسكان المستخدمين المؤهلين للمنتج. تنتج تقدير السكان الرئيسي.

32.2. لوحة مراقب الدولة

تمنع البلدان الصغيرة التي لا يمكنها توفير عينات كافية في لوحة السكان من أن تصبح غير مرئية تمامًا. لا تكفي بمفردها لإنتاج درجة البلد.

32.3. عينة إنصاف اللغة

يقيس أيضًا اللغات التي تعاني من نقص الموارد أو التي تقل شعبيتها عن المتوسطات السكانية. يعيد الاتصال بالنتيجة العالمية الرئيسية باستخدام الأوزان السكانية الحقيقية.

32.4. لوحة الوصول

يقيس احتياجات الوصول ومستخدمي تقنيات المساعدة التي قد يستثنيها اللوحة الرقمية القياسية. ملاحظات من هذه اللوحات الأربعة:

- guol,
- وزن،

- عالم الاستدلال

يجب تصنيفه بشكل منفصل من حيث الصيانة. لا يمكن استخدام نفس الملاحظة بالوزن الكامل في أكثر من لوحة واحدة دون تفسير.

33. لوحة البالغين الأولية

قاعدة المرشح الأساسية — القاعدة التأسيسية للمرشح

في الإصدار العام الأول من GEO-1000، لوحات معلومات المستخدم الحقيقي:

من المشاركين البالغين الذين يستوفون شروط المنتج والبلد

يوصى بإنشائه. هذا النهج:

- الموافقة،
- حماية البيانات،
- جمع لقطات الشاشة،
- البحث عالي المخاطر،
- موافقة الوالدين

يمكن أن يقلل من تعقيده. ومع ذلك، لا يمكن استنتاج الاستنتاج التالي: 'تم قياس تجربة الأطفال والمستخدمين الشباب باستخدام عينة البالغين.' في المنتجات التي تؤثر على الأطفال أو الشباب، بشكل منفصل:

- الموافقة الأخلاقية،
- البروتوكول المناسب للعمر،
- عمليات الوالد أو الوصي،
- حماية من الأذى

مطلوب.

34. لا يمكن تسمية أي إطار سكاني بـ "كل البشرية"

منتج:

- إذا لم يكن موجودًا في بعض الدول،
- إذا لم يدعم بعض اللغات،
- إذا كان يعمل فقط على أجهزة معينة،
- إذا كان لديه شروط عمرية أو خطط معينة،

النتيجة: لا يمكن تقديمه باعتباره "تجربة GEO لجميع الناس في العالم". التعبير الصحيح:

التمثيل المقدر في السكان المستهدفين المناسب للمنتج المحدد

يجب استخدامه. كلمة عالمي:

- متعدد الدول،
- متعدد اللغات,
- عدد كبير من السكان

يمكن أن يحدد قياسًا. لا يعني تغطية غير محدودة وعالمية.

35. الحد الأخلاقي الأساسي للسكان

تحديد السكان للقياس لا يعني اعتبار بعض الأشخاص بلا قيمة. قد تتلقى مجموعة وزنًا منخفضًا في النتيجة الرئيسية. ومع ذلك:

- ارتفاع المخاطر،
- العدالة اللغوية،
- سهولة الوصول،
- رؤية الدولة

قد يتم أيضًا فحصها بشكل منفصل. قد يتم استبعاد مجموعة من نتيجة نطاق الوصول الطبيعي لأنها لا تستطيع الوصول إلى منتج ما. يمكن أيضًا الإبلاغ عن هذا النقص في الوصول كفجوة تغطية عالمية. الغرض من NOMOS:

- ضغط الجميع في متوسط اصطناعي واحد،
- رؤية أكبر التجمعات السكانية فقط،
- يبدو أنه يرمز بشكل رمزي إلى المجتمعات الصغيرة بشخص واحد

ليس الهدف. الهدف هو:

> لتمثيل كل شخص في المقام الصحيح للسؤال الصحيح، دون خلط وزنه ورؤيته.

36. الأحكام الإلزامية المعيارية

CH05-N01

يجب تحديد سكان المستخدمين المستهدفين بشكل منفصل لكل منتج ذكاء اصطناعي.

CH05-N02

لا يمكن استخدام إجمالي عدد سكان الدولة مباشرة كمجموع المستخدمين المؤهلين للذكاء الاصطناعي.

CH05-N03

يجب ربط الفئة المستهدفة بالمنتج والدولة والعمر والوصول والواجهة والخطة واللغة وصياغة المُدخل والشروط الزمنية.

CH05-N04

يجب عمل نسخ من السكان المستهدفين وقواعد الاستبعاد قبل رؤية النتائج.

CH05-N05

لا يمكن نقل السكان المستهدفين لمنتج ذكاء اصطناعي إلى منتج ذكاء اصطناعي آخر دون تقديم شرح.

CH05-N06

عدم وجود وصول للمنتج لا يمكن تسجيله كنتيجة لتمثيل خاطئ؛ يجب الإبلاغ عنه بشكل منفصل كتغطية الوصول.

CH05-N07

لا يمكن طلب من المشاركين تجاوز قيود الوصول للمنتج أو الدولة.

CH05-N08

لا يمكن إضافة الوصول غير المدعوم رسميًا تلقائيًا إلى السكان الرئيسيين نطاق الوصول الطبيعي.

CH05-N09

يجب فصل وحدات الأفراد والحسابات والجلسات عن بعضها البعض.

CH05-N10

لا يمكن لحسابات متعددة أو لغات الشخص أن تضاعف وزن الفرد في السكان.

CH05-N11

لا يمكن تحديد أهلية اللغة بناءً فقط على لغة الدولة الرسمية.

CH05-N12

يجب تسجيل اللغة المطلوبة، ولغة الواجهة، ولغة الاستجابة، ولغة إتقان المستخدم بشكل منفصل.

CH05-N13

لا يمكن أن يتجاوز الوزن السكاني الإجمالي لشخص يستخدم عدة لغات شخصًا واحدًا بدون توضيح.

CH05-N14

يجب فصل اللوحة الموزونة بالسكان ولوحة مراقبي البلد.

CH05-N15

لا يمكن نشر ملاحظة واحدة أو عدة ملاحظات لمراقب البلد كنتيجة لدرجات البلد.

CH05-N16

في لوحة العدالة اللغوية، يجب إعادة الملاحظات التي تم أخذ عينات زائدة لها إلى وزنها السكاني الحقيقي في النتيجة العالمية.

CH05-N17

لا يمكن استبعاد المستخدمين الذين لديهم احتياجات إمكانية الوصول من الفئة المستهدفة لأن أداة جمع البيانات غير مناسبة.

CH05-N18

يجب الإبلاغ عن السكان المستهدفين، وإطار العينة، والسكان الذين تم الاتصال بهم، والمشاركين، والملاحظات التحليلية الصالحة بشكل منفصل.

CH05-N19

لا يجب تمثيل لوحة غير احتمالية أو تطوعية كمثال عشوائي للسكان العالميين.

CH05-N20

لا يمكن للوزن أن يجعل مجموعة غائبة تمامًا في إطار العينة قابلة للرصد.

CH05-N21

لا يمكن استخدام عدد السكان المحتملين المناسبين وعدد المستخدمين النشطين المعرضين كمعيار واحد.

CH05-N22

يجب فصل النتائج نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة عن بعضها البعض.

CH05-N23

لا يمكن استخدام نفس الفئة المستهدفة تلقائيًا لصياغات مُدخل الهوية والخدمة والسعر والقانون والتوصية.

CH05-N24

يجب أن تحمل بيانات السكان وحالة الوصول إلى المنتج وتوزيع اللغة طابعًا زمنيًا وإصدار المصدر.

CH05-N25

لا يمكن ضرب نسب المنفعة الحدية بشكل أعمى دون توضيح افتراض الاستقلالية.

CH05-N26

إذا لم يكن نوع المنتج مناسبًا للجمهور العام، فلا يمكن تطبيق طريقة السكان البشرية GEO-1000 بدون شرح.

CH05-N27

لا يمكن تضمين الأطفال أو المستخدمين القاصرين في اللوحة الرئيسية دون بروتوكول أخلاقي وقانوني منفصل.

CH05-N28

يجب نشر تغطية الوصول إلى المنتج ودقة التمثيل كنتائج منفصلة.

CH05-N29

لا يمكن للمصطلح العالمي جعل الدول أو اللغات أو الأعمار أو الواجهات أو الخطط التي تقع خارج النطاق غير مرئية.

CH05-N30

يجب أن يكون لسجل السكان المستهدف مالك بشري أو مؤسسي مسؤول.

37. أشكال الفشل

CH05-F01— احسب إجمالي السكان كالسكان المؤهلين

يتم تضمين سكان الدولة مباشرة في المقام دون تقييم المنتج أو العمر أو الإنترنت أو اللغة أو شروط الواجهة.

CH05-F02— طبق نفس خريطة العالم على جميع المنتجات

يتم قياس منتجات الذكاء الاصطناعي المختلفة وفقًا لنفس السكان المستهدفين.

CH05-F03— احسب المستخدم الذي يتجاوز الحد كمستخدم طبيعي

يتم تقديم الملاحظات التي تتجاوز شروط الوصول الرسمي كتجربة لسكان الدولة.

CH05-F04— احتساب غياب المنتج كإجابة خاطئة

الدول التي لا يمكن الوصول إليها تُعتبر صفرًا في دقة التمثيل.

CH05-F05— إخفاء غياب المنتج تمامًا

يتم إزالة السكان الكبار الذين لا يمكن الوصول إليهم من مقام نطاق الوصول الطبيعي؛ كما أن تغطية الوصول غير معروضة أيضًا.

CH05-F06— احتساب شخص واحد كأكثر من شخص واحد

حسابات الشخص نفسه أو لغته أو ملاحظات منتجاته تضاعف وزن السكان.

CH05-F07— احتساب الحساب كشخص

لا يُعتبر امتلاك حسابات متعددة أو استخدام الحساب المشترك.

CH05-F08— احتساب اللغة الرسمية كلغة مستخدم

تعتبر لغة الدولة الرسمية لغة التدقيق لجميع المشاركين.

CH05-F09— احتساب الأفراد متعددي اللغات مرتين

يحصل الفرد نفسه على الوزن الكامل للسكان في كل لغة.

CH05-F10— تعيين شخص واحد لبلد صغير وشرح درجة البلد

يتم تحويل ملاحظة مراقب البلد إلى تقدير إحصائي للبلد.

CH05-F11— حذف الدول الصغيرة تمامًا

نظرًا لأن حصة السكان منخفضة، فإن البلد أو اللغة غير مدرجة في أي لوحة تغطية إضافية.

CH05-F12— المبالغة في أمثلة عدالة اللغة في النتيجة الرئيسية

اللغات المفرط أخذ عينات منها لا تُعاد إلى وزنها السكاني الحقيقي.

CH05-F13— احتساب المستخدم النشط كسكان محتملين

يتم عرض بيانات الاستخدام الفعلية كما لو كانت تشمل جميع الأشخاص الذين يمكنهم الوصول إلى المنتج.

CH05-F14— احتساب السكان المحتملين كتعرض فعلي

يفترض أن استجابة كل من يمكنه استخدام المنتج تُرى في الواقع.

CH05-F15— احتساب اللوحة الطوعية كمثال عشوائي للعالم

يتم إخفاء الانحياز الناتج عن الاختيار الذاتي والانحياز في اللوحة.

CH05-F16— تجاهل خطأ التغطية باستخدام الترجيح

يُزعم أن المجموعات غير الموجودة في الإطار ممثلة فقط من خلال الترجيح.

CH05-F17— إزالة المستخدمين المستبعدين من أداة جمع البيانات من السكان

يتم تفسير مشاكل الوصول أو الجهاز كما لو أن المستخدم غير مناسب.

CH05-F18— تطبيق نفس السكان على جميع صياغات المُدخل

الهوية العالمية، التسعير المحلي، وأسئلة الاستشارات القانونية يتم دمجها على نفس المقام.

CH05-F19— تطبيق السكان العامين للمستهلك على منتج API

يتم نقل لوحة المستخدم البشري مباشرة إلى المطور أو منتج التطبيق.

CH05-F20— دمج الخطط المجانية والمدفوعة سرًا

يتم نشر توزيع منتج واحد دون تسجيل فرق الخطة.

CH05-F21— توحيد سكان الويب والموبايل

تظل اختلافات الوصول إلى الواجهة والسلوك غير مرئية.

CH05-F22— تجاهل الحد العمري

يتم إضافة فئات عمرية غير مناسبة إلى السكان المستهدفين أو اللوحة دون أي تفسير.

CH05-F23— ضرب المعدلات الحدية كما لو كانت مستقلة

يتم تجاهل العلاقة بين العمر والإنترنت واللغة والوصول.

CH05-F24— تغيير السكان بعد النتيجة

يتم إزالة بلد أو مجموعة مستخدمين منخفضة النتائج من السكان المستهدفين.

CH05-F25— استخدام خريطة وصول المنتج القديمة

لم يتم التحقق من الوصول الفعلي للمنتج في تاريخ القياس.

CH05-F26— خلط نتائج قاعدة المقارنة المشتركة و نطاق الوصول الطبيعي

تُعرض المقارنات المبنية على عدد السكان المشتركين كما لو أنها تعكس الوصول الفعلي للمنتج على مستوى العالم.

CH05-F27— تضييق السوق المستهدفة وفقًا للمصلحة التجارية

تعلن المنظمة فقط عن البلدان التي تحقق فيها النجاح كسوق مستهدفة.

CH05-F28— احتساب نطاق الوصول كدقة جغرافية
يُعرض التواجد في العديد من البلدان كدليل على التمثيل الدقيق.
CH05-F29— بيع الوصول المحدود كنجاح جغرافي مرتفع
يُعرض التميز العالي في عدد محدود من البلدان كسيادة عالمية.
CH05-F30— استخدام مصطلح "العالم بأسره" بدون سجل التغطية
استثناءات المنتج واللغة والبلد والخطة مخفية.

38. إجراء التدقيق

الخطوة 1— تثبيت منتج الذكاء الاصطناعي و واجهة المستخدم

تم تسجيل نوع الحساب وخطة واجهة المنتج والطراز المعروض وتاريخ القياس.

الخطوة 2— تصنيف سؤال القياس

صياغة المُدخل:

- الهوية العالمية،
- الخدمة المحلية،
- السعر،
- القانون،
- التوصية،
- الخطر العالي

إلى أي من عائلاتهم ينتمي؟

الخطوة 3— تحديد السكان المرجعيين العالميين

يتم تعريف الكون البشري الأولي الذي سيتم ربط النتيجة به.

الخطوة 4— تحديد العمر وأهلية البحث

يتم تسجيل شرط عمر المنتج، شرط عمر البحث، وشروط الموافقة ومعالجة البيانات.

الخطوة 5— التحقق من الوصول إلى المنتج بناءً على البلد

يتم تعيين حالة الوصول بين PA-0 و PA-5 لكل بلد.

الخطوة 6— تحديد التوافق مع الإنترنت والجهاز والواجهة

يتم تقدير السكان الذين يمكنهم استخدام الواجهة المقاس.

الخطوة 7— إنشاء عالم اللغة والمحلية

يتم فصل لغة المُدخل، ولغة الواجهة، ولغة الاستجابة، وكفاءة المستخدم في اللغة.

الخطوة 8— تطبيق صلة المُدخل

يتم تقييم ما إذا كان جميع المستخدمين ينتمون إلى الفئة المستهدفة لعائلة المُدخل ذات الصلة.

الخطوة 9— اختيار نموذج السكان المناسب لنوع المنتج

يتم فصل المنتج إلى مستهلك، أو مؤسسة، أو تعليم، أو API، أو منتج مضمّن.

الخطوة 10— مقارنة السكان المستهدفين بإطار العينة

يتم تسجيل ذلك كعدم تغطية أو تغطية زائدة.

الخطوة 11— تحديد اللوحات الرئيسية والمكمّلة

مجالات لوحة السكان، ولوحة مراقب الدولة، ولوحة عدالة اللغة، ولوحة الوصولية مرقمة بالإصدارات.

الخطوة 12— فصل العوالم نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة

يتم تسجيل نتائج المنتج داخل المنتج ومجموعات المقارنة بين المنتجات بشكل منفصل.

الخطوة 13— فصل نتائج نطاق الوصول والدقة

يتم الإبلاغ عن الدقة بين السكان الذين يمكن للمنتج الوصول إليهم والمستخدمين الذين تم الوصول إليهم بشكل منفصل.

الخطوة 14— تجميد ملف السكان

مصادر البيانات والافتراضات والتاريخ وإصدار السكان مثبّتة.

الخطوة 15— تتبع التغييرات المادية

إذا تغير الوصول إلى المنتج أو الشروط أثناء القياس، نسخة جديدة يُنشأ.

39. الأدلة المطلوبة

هوية منتج الذكاء الاصطناعي؛ واجهة المستخدم؛ شروط الخطة والحساب؛ تاريخ القياس؛ البلدان التي يمكن الوصول فيها إلى المنتج؛ حالة الوصول على مستوى البلد؛ شروط العمر؛ شروط المشاركة القانونية والأخلاقية؛ سجلات السكان حسب البلد؛ توزيعات العمر؛ سجلات الوصول إلى الإنترنت والأجهزة؛ توزيعات اللغة والموقع؛ دعم اللغة للمنتج؛ إمكانية الوصول للواجهة؛ عائلة الطلب المستهدفة؛ تعريف السكان المستهدفين؛ إطار العينات؛ نقص التغطية؛ التغطية الزائدة؛ نوع اللوحة؛ طريقة اختيار المشاركين؛ حالة اللوحة حسب الاحتمالية أو الاشتراك الاختياري؛ أوزان المستخدمين متعدد اللغات؛ تغطية مراقب البلد؛ تغطية عدالة اللغة؛ تغطية لوحة الوصول؛ عدد السكان نطاق الوصول الطبيعي. سكان قاعدة المقارنة المشتركة؛ بيانات المستخدم النشط، ومكان استخدامها؛ حساب تغطية الوصول؛ الافتراضات وتحليلات الحساسية؛ نسخة السكان؛ وتاريخ التحقق الأخير.

الشخص أو المؤسسة المسؤولة

40. قائمة تدقيق التدقيق

أي منتج ذكاء اصطناعي و واجهة المستخدم يتم قياسه؟ لأي تاريخ تم تحديد العدد المستهدف للسكان؟ هل تم فصل إجمالي سكان البلد والسكان المؤهلين للاستخدام؟ هل تم تطبيق متطلبات عمرية؟ هل تم تقييم الوصول إلى الإنترنت والجهاز؟ هل المنتج متاح رسميًا على أساس كل بلد؟ هل طُلب من المشاركين تجاوز قيود الوصول؟ هل تم فصل الخطط المجانية والمدفوعة؟ هل تم فصل واجهات الويب والجوال؟ هل غيرت مجموعة صياغات المُدخل السكان المستهدفين؟ كيف تم تحديد ملاءمة اللغة والمنطقة؟ هل تم احتساب المستخدمين متعددين اللغات مرتين؟ هل الفرق بين السكان المستهدفين وإطار العينة واضح؟ هل اللوحة احتمالية أم تطوعية؟ هل تم قياس خطأ التغطية؟ هل تم استبعاد المستخدمين ذوي الاحتياجات من الوصول بصمت؟

هل تم تحويل ملاحظات مراقب البلد إلى درجة للبلد؟ هل كانت أمثلة عدالة اللغة لها وزن مفرط في الدرجة الرئيسية؟ هل تم فصل نتائج نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة؟ هل التغطية ودقة التمثيل منفصلتان؟ هل تم فصل نتائج المستخدمين النشطين والمستخدمين المحتملين؟ هل نوع المنتج مناسب للوحة السكان البشرية؟ هل تم ضرب معدلات الهامش في أوزان السكان بشكل أعمى؟ هل تم تغيير تعريف السكان بعد رؤية النتائج؟ هل النسخة المستخدمة من بيانات السكان والوصول موثقة بالإصدار؟ هل المجموعات غير المختبرة مرئية؟ هل مصطلح "عالمي" متسق مع التغطية الفعلية؟ هل هناك مالك مسؤول واضح لسجل السكان؟

41. الاعتراضات والردود

اعتراض —1 "أليس من الأنصاف التوزيع مباشرة وفقًا للسكان العالمين؟"

يعتبر عدد سكان العالم قاعدة بداية قوية. ومع ذلك، فإن تضمين الأشخاص الذين لا يستطيعون استخدام المنتج في العينة الرئيسية للمنتج لا يولد تجربة مستخدم حقيقية. الطريقة الصحيحة هي:

- وزن السكان ضمن السكان المؤهلين للمنتج،
- تغطية الوصول العالمي.

يجب الإبلاغ عن كلا النتيجتين.

الاعتراض —2 "أليس من المفترض أن يأتي معظم المستخدمين من الدول الكبيرة؟"

يجب أن يكون للدول التي تفي بشروط المنتج واللغة والوصول حصة كبيرة بطبيعة الحال. ومع ذلك، لمجرد أن إجمالي السكان كبير، لا يمكن توليد تجربة مستخدم مزيفة من دولة لا يمكن الوصول فيها إلى المنتج.

الاعتراض —3 "لماذا سيكون خطأً أن نعطي شخصًا واحدًا لبلد صغير مثل سان مارينو؟"

الشخص ذو قيمة للاستكشاف ومراقبة الدولة. الخطأ هو تقديم نتائج هذا الشخص كـ 'درجة الدولة'. يمكن حماية رؤية البلد الصغير من خلال عينة مراقبي الدولة. يظل الوزن السكاني الرئيسي وفقًا لحصة السكان الفعلية.

الاعتراض —4 "إذا أعطينا شخصًا واحدًا على الأقل لكل دولة، ألن يبدو أننا أكثر عالمية؟"

يتم توفير تغطية جغرافية أوسع. ومع ذلك، قد يتم تشويه تقدير السكان. لذلك:

- النطاق الجغرافي،
- الكثافة السكانية

إنها نتائج منفصلة. قد يكون لوحة نطاق لتخصيص شخص واحد لكل دولة. إنها ليست لوحة سكانية.

اعتراض —5 "إذا أعطينا صفرًا لمنتج في دولة غير قابلة للوصول، ألن يكون المنتج معاقبًا بشكل غير عادل؟"

لا تُحسب الدولة غير القابلة للوصول كصفر في دقة التمثيل. بل تصبح مرئية في مقياس تغطية الوصول. هذا التمييز يحمي المنتج من كل من العقوبات غير العادلة والشمولية الزائفة.

اعتراض —6 "إذا كانت هناك بيانات استخدام حقيقية، فلماذا نحتاج بيانات سكانية؟"

تُظهر بيانات الاستخدام الحقيقية التأثير التشغيلي. ومع ذلك، فإن المنتج:

- أي الفئات السكانية التي لم يصل إليها.
- عدم المساواة الجغرافية في قاعدة المستخدمين،
- العدالة المحتملة للمستخدم

لا يظهر بمفرده. النوعان المختلفان من البيانات يجيبان على أسئلة مختلفة.

اعتراض —7 'نظرًا لأن جميع المستخدمين سيطرحون نفس صياغة المُدخل، لماذا تعتبر أهمية صياغة المُدخل مهمة؟'

قد تكون أسئلة الهوية ذات معنى في مجموعة سكانية واسعة. أما أسئلة السعر أو القانونية أو الاستشارية فهي ذات معنى فقط لبعض المستخدمين والأسواق. قد تشوه الفئة السكانية غير ذات الصلة توزيع الإجابات بعيدًا عن واقع المستخدم.

اعتراض —8 "إذا كان الشخص يعرف لغات متعددة، فلماذا لا يتم احتسابه بالكامل في كل اللغات؟"

يمكنك اختبار جميع اللغات في المقارنة عبر اللغات. ومع ذلك، فإن مضاعفة نفس الشخص في وزن السكان تزيد السكان العالميين بشكل مصطنع. يجب تقسيم وزن الشخص، أو استخدام لوحة مطابقة منفصلة.

اعتراض —9 "ألا يجعل اختبار البالغين فقط النتيجة العالمية غير مكتملة؟"

يفعل ذلك. يجب توضيح هذا القيد بوضوح. البدء بلوحة للبالغين قد يكون خيارًا أخلاقيًا وعمليًا. هذا لا يعتبر حكمًا على المستخدمين الأطفال.

اعتراض —10 'أليس مجلس البحث عبر الإنترنت كافيًا؟'

يمكن أن يكون ذا قيمة كبيرة للبحث التجريبي والبحث الموزون. لكن المجلس:

- كم من السكان المستهدفين يغطي،
- من يتم استبعاده،
- كيف يؤدي المشاركة الطوعية إلى التحيز

يجب توضيح ذلك. استخدام عينة ليس خطأً. عرض العينة كما لو كانت عشوائية مثل سكان العالم هو خطأ.

اعتراض —11 "لماذا عينة الوصولية مرتبطة بالدرجة الرئيسية GEO؟"

لأن جزءًا من مستخدمي الذكاء الاصطناعي الحقيقيين يستخدمون تكنولوجيا مساعدة أو أشكال مختلفة من التفاعل. إذا استبعدت أداة جمع البيانات هؤلاء الأشخاص، قد لا تمثل الدرجة الرئيسية كامل الفئة المستهدفة.

اعتراض —12 "ألا يجعل وجود العديد من تعريفات السكان المختلفة من المستحيل الحصول على درجة واحدة NOMOS؟"

لا. يمكن إنتاج نتيجة واحدة فقط. ومع ذلك، يجب أن يكون من الواضح أي سؤال سكاني تجيب عليه. على سبيل المثال:

- نتيجة السكان نطاق الوصول الطبيعي
- النتيجة قاعدة المقارنة المشتركة
- النتيجة ذات الصلة بالسوق
- نتيجة العدالة اللغوية

يمكن أن تحمل سجلات منفصلة. تنشأ دلالة الرقم من وضوح المقام الخاص به.

44 الحكم العام للفصل

GEO-1000 يحتوي على '1,000' في اسمه، لكن قيمة هؤلاء الأشخاص 1,000 لا تنشأ من الرقم وحده. بل تنشأ من إجابات هذه الأسئلة: من هم؟ من أي سكان تم اختيارهم؟ هل يمكنهم فعليًا استخدام المنتج؟ أين يعيشون؟ أي اللغات يستخدمون؟ كم مرة تم عد نفس الشخص؟ هل الدول الصغيرة مرئية؟ هل تحتفظ الدول الكبيرة بوزن سكانها الحقيقي؟ أين يأخذ التقييم في الاعتبار الأشخاص الذين لا يمكن للمنتج الوصول إليهم؟ من استبعدته بنية العينة؟ أي نتيجة تقيس الوصول المحتمل، وأي نتيجة تقيس التعرض الفعلي؟ قد يكون لدولة كبيرة عدد كبير من السكان

المؤهلين للمنتج وبالتالي تستحق وزناً كبيراً في العينة الرئيسية. ولكن إذا لم يكن المنتج متوفراً هناك، يجب على التدقيق عدم اختلاق مستخدمين. ويجب الإبلاغ عن فجوة الوصول بشكل منفصل.

إذا كانت الدولة صغيرة جدًا، يجب أن يظل الوزن السكاني الرئيسي منخفضًا. ومع ذلك، لا يجب أن يكون غير مرئي تمامًا. يمكن ملاحظته من خلال لوحة مراقبة. إذا كانت اللغة بها عدد سكان قليل، فقد لا تحدد المتوسط العالمي. ومع ذلك، إذا كان هناك تحريف شديد في تمثيل تلك اللغة، يجب أن يكون مرئيًا أيضًا. إذا كان الشخص يعرف ثلاث لغات، يمكنه إنتاج ثلاث تجارب مختلفة. ومع ذلك، هم ليسوا ثلاثة أشخاص مختلفين. إذا كان منتج الذكاء الاصطناعي يقدم سلوكًا معينًا فقط في خطة مدفوعة، فإن درجة المستخدمين المجانيين لا تمثل تلك الخطة. لوحة السكان البشرية مناسبة لمنتج محادثة استهلاكية. قد لا يكون نفس نموذج السكان مناسبًا لواجهة برمجة التطبيقات أو منتج المؤسسة. لذلك، فإن قانون القياس الخامس لـ NOMOS هو كما يلي:

> حتى أعلى الدرجات حساسية تعطي الإجابة الصحيحة على السؤال الخطأ إذا كان المقام غير صحيح.

قانونه السادس هو:

> وزن السكان لا يحل محل وضوح البلد واللغة.

أما القانون السابع فهو كما يلي:

> الشخص الذي لا يمكنه الوصول إلى [شيء ما] لم يتم تمثيله بشكل خاطئ؛ ومع ذلك، فإنهم جزء من حدود الوصول العالمية.

قانونها الثامن هو كما يلي:

> وجود شخص لديه عدة حسابات أو لغات أو جلسات لا يزيد من عدد السكان البشر.

القانون التاسع هو كما يلي:

> القياس العالمي ليس ادعاءً بأن الجميع يتم قياسهم؛ إنه سجل صادق لمن تم قياسه ومن لم يتم.

طلب قسم 5 من NOMOS

> لا تعطيني فقط أعداد سكان الدول. / أظهر أي الأشخاص يمكنهم فعليًا استخدام المنتج.
>
> لا تملأ دولة كبيرة بمستخدمين مزيفين عندما لا يكون المنتج موجودًا هناك.
>
> لا تعطي شخصًا واحدًا لدولة صغيرة وتعلن أن رد هذا الشخص هو صوت الدولة.
>
> لا تكرر شخصًا بقدر حساباته ولغاته وجلساته.
>
> لا تخلط بين اللغة الرسمية والاستخدام الفعلي للغة من قبل الناس.
>
> لا تجعل مستخدمًا مجانيًا ممثلًا لخطة مدفوعة، أو مستخدمًا للهواتف المحمولة ممثلًا لواجهة الويب.
>
> لا تعتبر لوحة المتطوعين سكانًا عشوائيين من العالم.
>
> لا تمحُ مستخدمًا ذو احتياجات وصول من السكان لأن الأداة لا تعمل.
>
> لا تزيل السكان الذين لا يصلهم المنتج من درجة الدقة وتخفي فجوة الوصول العالمية.
>
> لا تطبّق نفس الحصة على سؤال السعر كما تطبّق على سؤال الهوية، ولا على علامة تجارية عالمية كما تطبّق على ترخيص محلي.
>
> أولاً، حدد المنتج. / ثم حدد السؤال. / ثم حدد الكون البشري. / ثم قيد الوصول، اللغة، العمر، والواجهة. / ثم اشرح من لا يمكنك اختياره. / وفقط بعد ذلك وزع أشخاص *1,000.*

جملة ختام الفصل

في *GEO-1000*، العدالة ليست في إعطاء نفس عدد الأشخاص لكل دولة؛ بل هي تمثيل كل فرد في السكان المستهدفين الصحيحين، بالترجيح المناسب، دون إخفاء الأماكن التي لم يتم قياسهم فيها.

النواة المعيارية

يجب على كل تدقيق سكاني في **GEO** أن يحدد، قبل فحص النتائج، مجتمعًا مستهدفًا من المستخدمين مقيدًا بزمن ومحددًا لكل منتج، وذلك لكل من: - منتج الذكاء الاصطناعي وواجهة الاستخدام، - البلد أو الاختصاص القضائي، - العمر وشروط الوصول المشروع، - حالة الحساب أو الخطة، - اللغة والإعداد المحلي، - الواجهة، - عائلة المُدخلات، - وغرض القياس. ولا يجوز اعتبار إجمالي السكان المقيمين مجتمعًا مؤهلًا لاستخدام المنتج من دون تعديلات مبررة للأهلية. ويجب الإبلاغ عن عدم توافر المنتج بوصفه قيدًا على تغطية الوصول، لا أن يُحتسب بصمت تمثيلًا غير صحيح أو يُحذف بصمت من النطاق العالمي. ويجب أن تظل مفاهيم إمكانية الوصول، والاستخدام النشط، والتعرض الفعلي، ونطاق الوصول الطبيعي، وقاعدة المقارنة المشتركة، والأهمية السوقية، وقابلية الرصد البحثي منفصلة. ولا يجوز التعامل مع الأشخاص والحسابات والجلسات واللغات وملاحظات المنتجات بوصفها وحدات سكانية قابلة للتبادل. ويجب إعادة ترجيح البلدان واللغات ومجموعات إمكانية الوصول المفرط في أخذ عينات منها عند تقدير المجتمع، مع إبقائها ظاهرة بصورة منفصلة لتحليل العدالة والتغطية. ويجوز لملاحظات مراقب الدولة أن تثبت وجود تغطية للرصد الجغرافي، لكنها لا يجوز أن تُقدَّم بوصفها درجات موثوقة إحصائيًا على مستوى الدولة. ولا يجوز إلزام المشاركين بتجاوز قيود المنتج أو الموقع الجغرافي أو العمر أو الهوية أو الحساب للانضمام إلى اللوحة السكانية الرسمية. ويجب أن يكون لكل إطار سكاني رقم إصدار ومصادر موثقة، وأن يراعي عدم اليقين، وأن يُنسب إلى شخص أو منظمة مسؤولة.

عينة GEO-1000 المتناسبة مع السكان

الأسس المنهجية / K04؛ K05؛ K07؛ K08

حد الفصل

أقرت المادة 5 الحكم التالي:

> يجب تعريف جمهور المستخدمين المستهدف لكل منتج ذكاء اصطناعي بشكل منفصل. إجمالي سكان أي دولة ليس بالضرورة جمهور المستخدمين المؤهلين للمنتج في تلك الدولة.

والآن نأتي إلى السؤال التالي:

> من جمهور المستخدمين المؤهلين المحدد لـ *GEO-1000*، كم عدد الأشخاص الذين سيتم اختيارهم، من أي دولة، وما هي اللغة، وبأي طريقة؟

جمهور المستخدمين المؤهلين للمنتج الذكاء الاصطناعي هو:

- 40% في البلد أ،
- 25% في البلد ب،
- 15% في البلد ج،
- 0.04% في بلد صغير

إذا كان متاحًا، كيف يجب توزيع الملاحظات الصالحة 1,000؟ كيف سيتم تحويل الحصص الكسرية إلى أعداد صحيحة؟ هل سيتم استبعاد البلدان التي تكون حصة سكانها أقل من شخص واحد تمامًا؟ هل ستتغير أوزان السكان عند منح شخص واحد على الأقل لكل بلد؟ هل يمكن تجميع البلدان الصغيرة في مجموعة منفصلة؟ إذا تم أخذ عينات مفرطة لبعض اللغات لضمان الإنصاف اللغوي، كيف سيتم حساب النتيجة العالمية؟ إذا كان الهدف هو الحصول على 100 ملاحظات صالحة في بلد ما، كم عدد الأشخاص الذين يجب دعوتهم؟ من سيحل محل المشارك إذا لم يكمل المهمة؟ هل يمكن استبدال المستخدم الذي يقدم إجابات خاطئة؟ إذا اختبر نفس الشخص أكثر من منتج ذكاء اصطناعي، هل ستظل العينة مستقلة؟

حتى مع عينة خام من 1,000 شخصًا، هل يمكن للأوزان غير المتساوية بدرجة كبيرة أن تقلل المعلومات الإحصائية الفعالة إلى ما هو أدنى من 1,000 ملاحظات مستقلة ومتساوية الوزن؟ هذه ليست مجرد تفاصيل تشغيلية. تصميم العينة يحدد مقدار التأثير الذي يمتلكه كل شخص على درجة GEO. قد تؤدي العينة المعيبة إلى:

- إلى إعطاء وزن خاطئ للإجابات الصحيحة،
- إلى تمثيل رمزي مفرط للدول الصغيرة،
- إلى جعل السكان الكبيرين يبدو أصغر مما هم عليه،
- إلى جعل اللغات منخفضة الموارد غير مرئية،
- إلى عرض مستخدمي التكنولوجيا الطوعية كما لو كانوا يمثلون سكان العالم.

يمكن تعديلها لتحقيق الدرجة المرغوبة بعد مراجعة النتائج. يتطلب الهيكل التدقيقي الذي تم إنشاؤه مسبقًا أن يكون كل تدقيق قابلاً للقياس وأن يحمل النموذج/المنتج تاريخًا وبلدًا ولغة ومجموعة استفسارات وعدد التكرارات وسجل القياس. ينطبق هذا القسم بنفس الانضباط على تصميم العينة. نظرًا لأن الأدلة الحقيقية تتطلب تحديد الحدود والسياق والزمن؛ فإنها تتطلب أيضًا إظهار كيفية إنشاء العينة. يتناول هذا القسم:

- كيفية توزيع الملاحظات الصالحة 1,000 لهيئة السكان الرئيسية،
- كيفية إنشاء الشرائح والخلايا،
- كيفية إدارة الدول الصغيرة،
- كيفية تحويل الكسور إلى أعداد صحيحة.
- كيفية تحديد وزن الإفراط في العينة،
- كيفية فصل دعوات المشاركة، والنسخ الاحتياطي، وأهداف الملاحظات الصحيحة،
- كيفية وضع أوزان التصميم والتحليل،
- سبب اختلاف حجم العينة الفعال عن العدد الخام

يحدد. هذا الفصل لم يتطرق بعد إلى:

- من أي مزود لوحة سيتم اختيار المشاركين،
- التفاصيل الكاملة لتنفيذ الميدان في العينات الطوعية أو الاحتمالية،
- مراجعة وتقييم الإجابات،
- طريقة فترة الثقة النهائية،
- جميع صيغ تسجيل NOMOS

لم يتم الانتهاء منها بالكامل. مهمة القسم 6 هي:

> حوّل السكان المستهدفين المحددين إلى عينة *GEO-1000* مجمّدة، قابلة للوزن، وقابلة لإعادة الإنتاج قبل رؤية النتيجة.

التحدي **NOMOS**

أنت تقوم بإعداد عينة عالمية من أشخاص 1,000 لمنتج ذكاء اصطناعي. السكان المؤهلون للمنتج في عشرة دول هم كما يلي:

البلد	سكان المستخدمين المؤهلين
A	400,000,000
B	250,000,000
C	150,000,000
D	100,000,000
E	60,000,000
F	30,000,000
G	8,000,000
H	1,500,000
I	400,000
J	100,000
الإجمالي	1,000,000,000

الحصص الخام وفق نسبة السكان الصافية:

البلد	الحصة الخام
A	400.0
B	250.0
C	150.0
D	100.0
E	60.0
F	30.0
G	8.0
H	1.5
I	0.4
J	0.1

لا يمكن اختيار نصف شخص. لا يمكن إرسال 0.4 شخصًا إلى أي دولة. إذا قمت بتقريب الحصص الخام إلى أقرب عدد صحيح، فقد يتجاوز الإجمالي أو يقل عن 1,000. إذا أعطيت شخصًا واحدًا على الأقل لكل دولة، فإنك تُمثل الدولتين **I** و **J** بشكل مفرط مقارنة بحصصهما السكانية.

إعطاء شخص واحد للدولة **I** يزيد حصتها النظرية بمقدار 2.5 مرة. إعطاء شخص واحد للدولة **J** يزيد حصتها النظرية بمقدار 10 مرة. يمكن تصحيح هذا التمثيل المتطرف إلى حد كبير. ولكن لا يمكنك إنتاج درجة دولة مع شخص واحد. ماذا يحدث إذا لم تعطي أحدًا للدولتين **I** و **J**؟

إذا عرّفت هذه البلدان على أنها طبقات منفصلة وواجبة، فإن احتمال اختيار بعض الأشخاص في السكان المستهدفين يصبح صفراً. في هذه الحالة، يضعف الادعاء بوجود "عينة احتمالية تمثل كامل السكان العالميين المناسبين". الآن فكر في قرار أكثر خطورة. أولاً، تجمع الإجابات.

المستخدم الوحيد في البلد **J** يحصل على إجابة غير صحيحة. ثم تقول: "حجم العينة في هذا البلد صغير جداً، فلنستبعده من النطاق." في البلد A، هناك العديد من الإجابات الإيجابية. تحتفظ به بالوزن الكامل. هذا لم يعد تقريباً. إنه تغيير في العينة بناءً على النتائج.

في حالة أخرى، تقوم بزيادة معدل السجلات غير الصالحة في مجموعة لغوية ذات نتائج منخفضة، باستخدام المشاركين الاحتياطيين فقط حتى يتم الحصول على استجابة إيجابية. هذا أيضًا ليس عينة. إنه آلية اختيار استجابة. الحكم الأول في هذا القسم هو:

> يجب تأسيس العينة قبل رؤية استجابات الذكاء الاصطناعي.

ينص حكمه الثاني على:

> لا يمكن أن يعتمد احتمال اختيار مشارك على صحة إجابته.

تنص الفقرة الثالثة منه على:

> يجب إدارة وضوح الدولة الصغيرة ووزن السكان في النتيجة العالمية بشكل منفصل.

تنص المادة الرابعة على:

> يمكن إجراء أخذ عينات زائد؛ لكن زيادة الوزن ليست إلزامية.

تنص أحكامه الخامسة على:

> سجلات *1,000* الخام لا تحتوي دائمًا على معلومات إحصائية مكافئة لملاحظات *1,000*.

1. غرض الفصل

الغرض من هذا القسم هو تحويل الملاحظات الصالحة المستهدفة لـ 1,000 لكل منتج ذكاء اصطناعي وموجة قياس إلى تصميم عينة يمكن التحقق منه بناءً على عدد المستخدمين المؤهلين للمنتج. يقوم القسم بتوحيد التمييزات التالية:

- السكان المستهدفون مقابل العينة المختارة
- الحصة الخام مقابل التوزيع الصحيح للأعداد الصحيحة
- لوحة السكان الرئيسية مقابل اللوحات التكميلية
- رؤية البلد مقابل وزن البلد
- الطبقات مقابل الخلايا التجريبية
- طبقات البلد الكبير مقابل مجموعة البلد الصغير
- التوزيع النسبي مقابل العينات الزائدة غير المتناسبة
- حصة العينة مقابل حصة السكان
- عدد الدعوات مقابل الهدف من الملاحظات الصالحة
- المشارك الاحتياطي مقابل الاستبدال بعد النتائج
- عدم صلاحية جمع البيانات مقابل مخرجات الذكاء الاصطناعي السلبية
- وزن التصميم مقابل وزن النتائج
- حجم العينة الخام مقابل حجم العينة الفعّال
- عينة الدولة مقابل عينة اللغة
- المستخدم متعدد اللغات مقابل أكثر من شخص واحد
- الاختيار ذو المرحلة الواحدة مقابل الاختيار متعدد المراحل
- توزيع نطاق الوصول الطبيعي مقابل توزيع قاعدة المقارنة المشتركة
- تخصيص الموجة الفوري وتخصيص سلسلة الوقت
- تقدير السكان وتحليل التعرف على الدولة أو اللغة
- الملاحظات الرئيسية 1,000 مع المراقبين المساعدين وعينات العدالة

في نهاية هذا القسم، يجب أن يكون كل تدقيق قادراً على الإجابة على هذه الأسئلة:

> لماذا تم توزيع ملاحظات *1,000* بهذه الأعداد على هذه الدولة واللغة ومجموعات المستخدمين؟
>
> كيف تم تحديد وزن كل ملاحظة في النتيجة العالمية؟
>
> من كان لديه احتمال صفري للاختيار ولماذا؟
>
> كم عدد الأشخاص الذين تمت دعوتهم، وكم عدد الذين أكملوا، وكم عدد الملاحظات التي بقيت صالحة؟
>
> هل تم تجميد نتائج العينة قبل عرضها؟

2. الأحكام المعيارية المركزية

يجب أن يستند العينة السكانية الأساسية GEO-1000 إلى تصميم أخذ عينات طبقي يخصص احتمالات اختيار معروفة أو متوقعة للأشخاص في إطار السكان المؤهلين للمنتج المجمد مسبقًا لكل منتج ذكاء اصطناعي وموجة قياس. الهدف الافتراضي للوحة السكان الأساسية هو:

> على الأقل *1,000* ملاحظة صالحة لكل منتج ذكاء اصطناعي وموجة القياس الرئيسية

يتم الحفاظ عليه. هذا الرقم:

- ليس عدد الأشخاص المدعوين,
- ليست عدد الحسابات،
- ليس عدد لقطات الشاشة الخام،

ليست مجموع جميع اللوحات التكميلية. 1,000 هو الهدف للملاحظات الصالحة للمنتج التي تتوافق مع البروتوكول في لوحة السكان الرئيسية. يمكن لمراقبي الدول، لوحة عدالة اللغة، ولوحة الوصولية إنشاء عينات إضافية. هذه العينات الإضافية:

- يمكن أن تزيد من حجم الحقل الإجمالي فوق 1,000،
- يمكن استخدامها مع الأوزان المناسبة في نتيجة السكان العالميين،

ولكن لا يمكن استخدامها لتقليل هدف الملاحظات الصالحة 1,000 للوحة الرئيسية بشكل خفي.

3. التمثيل الأساسي للعينة

لمنتج AI معين a، موجة القياس w، وإطار السكان المستهدف، فإن الطبقات:

```
h∈ H_{a,w}
```

يتم الإشارة إليها. لكل طبقة:

- Nh: تقدير عدد السكان المناسب للمنتج في الطبقة
- N: إجمالي عدد السكان المستهدف
- Wh: وزن السكان للطبقة
- n: الهدف من ملاحظات اللوحة الأساسية الصالحة
- qh: الحصة النسبية الخام
- nh: العدد الصحيح المستهدف للملاحظات الصالحة المخصصة

إجمالي السكان المستهدف:

```
N= Σ_{h=1}^H N_h
```

وزن السكان للطبقة:

```
W_h= N_h/ N
```

الحصة النسبية الخام:

```
q_h= nW_h
```

هو. التخصيص النسبي الرئيسي، في ظل الظروف المثالية:

```
n_h≈ nW_h
```

يجب أن يكون. ومع ذلك:

- قد يكون qh كسرياً،
- بعض الخلايا قد تقل عن شخص واحد،
- قد تكون هناك حاجة لعينة تشخيصية دنيا لفئات اللغة والبلد الفرعية،
- يمكن أيضاً فحص المجموعات الصغيرة عالية المخاطر،

قد تختلف معدلات اكتمال الميدان. لذلك، الحصة الخام ليست التخصيص النهائي مباشرة.

4. ثلاثة أغراض مميزة لتصميم العينة

قد لا تفي العينة نفسها بجميع الأغراض بنفس القوة. يجب فصل ثلاثة أغراض.

4.1. تقدير السكان العالميين

يجيب على السؤال:

> ما هي النتيجة التمثيلية التي سيربها الشخص العادي في مجموعة المستخدمين العالمية المناسبة للمنتج؟

الوزن السكاني هو الأولوية الأساسية.

4.2. تشخيص الفئة الفرعية

يجيب على السؤال:

> ما الانحرافات التمثيلية التي تحدث في دول معينة أو لغات أو خطط أو مجموعات مستخدمين؟

قد يكون من الضروري أخذ عينات إضافية من المجموعات الصغيرة.

4.3. التغطية الجغرافية واللغوية

يجيب على السؤال:

> في أي دول ولغات يوجد على الأقل مراقبة فعلية للمستخدمين؟

تصبح لوحات مراقب الدولة وعدالة اللغة مهمة. لا يمكن دمج هذه الأهداف الثلاثة بشكل غير مفهوم في عينة واحدة.

5. هيكل العينة الرئيسي

يتكون هيكل العينة الرئيسي المرشح لـ GEO-1000 من أربع طبقات.

5.1. لوحة السكان

كل منتج وموجة ذكاء اصطناعي تحتوي على حد أدنى من 1,000 ملاحظات صالحة كلوحة رئيسية. وهي تنتج تقديراً للسكان العالميين.

5.2. تراكب مراقب الدولة

يوفر ملاحظات جغرافية إضافية للدول الصغيرة التي لا يمكن تمثيلها مباشرة في لوحة السكان الرئيسية. يمكن أن تشير هذه الملاحظات إلى:

- حدث واحد،
- حالة الوصول،
- احتياجات البحث المستقبلية

هم لا يولدون نتيجة لدولة بمفردهم.

5.3. زيادة عينة العدالة اللغوية

يُنتج ملاحظات إضافية للغات منخفضة الموارد أو اللغات غير الممثلة بشكل كافٍ في اللوحة الرئيسية. يتم إعادة وزنها وفقًا لأوزان السكان الفعلية في النتيجة العالمية.

5.4. زيادة عينة سهولة الوصول

يقيس أيضًا مستخدمي التكنولوجيا المساعدة وسهولة الوصول، الذين لا يتم تمثيلهم بشكل كافٍ من قبل اللوحة القياسية. يتم تسجيل الكون المستهدف ووزن هذه الملاحظات بشكل منفصل.

6. ما هو الشريحة؟

الطبقة السكانية هي شريحة من السكان يتم إدارتها بشكل منفصل من حيث تخصيص العينة واحتمالية الاختيار. يمكن أن تكون الطبقة مزيجًا من واحد أو أكثر من المجالات التالية:

- البلد
- المنطقة
- اللغة
- الموقع
- خطة الذكاء الاصطناعي
- الويب أو الجوال
- مسيطر عليه أو عينة المستخدمين في ظروف الاستخدام الطبيعية
- حالة الوصول
- نوع المستخدم
- الاختصاص القضائي

إذا تم جعل كل تركيبة صغيرة طبقة منفصلة، تصبح العينة مجزأة. على سبيل المثال: الجمع بين البلد × اللغة × الخطة × الجوال/الويب × الفئة العمرية × نوع المستخدم يمكن أن ينتج عنه مئات أو آلاف الخلايا. الملاحظات في 1,000 لا توفر تقييمات موثوقة لمعظم هذه الخلايا. لذلك، يجب أن تكون التجزئة:

- كبيرة.
- قابل للقياس،
- مناسب لحجم العينة

يجب أن تكون.

7. الطبقة ليست هي نفسها خلية التقرير

يمكن عرضها بشكل منفصل على بطاقة نتائج المجموعة. ومع ذلك، لا يجب أن تكون طبقة منفصلة في تصميم العينة. على سبيل المثال:

- لغة صغيرة،
- مجموعة وصول محددة،
- فئة عمرية معينة

يمكن اختياره ضمن طبقة أوسع في العينة الرئيسية. لاحقًا، يمكن الإبلاغ عنه بشكل منفصل في التحليل. والعكس ممكن أيضًا: يمكن فصل منطقتين بشكل طبقي من أجل كفاءة العينات. ومع ذلك، في التقرير العام، يمكن دمجهما تحت نتيجة إقليمية واحدة. يجب أن يحمل تصميم الطبقات وهيكل التقرير سجلات منفصلة.

8. تسلسل إنشاء الطبقة

تسلسل المرشح للوحة السكان الرئيسية كما يلي:

- منتج الذكاء الاصطناعي و واجهة المستخدم
- نطاق الوصول الطبيعي أو إطار السكان قاعدة المقارنة المشتركة
- الدولة أو مجموعة دول صغيرة
- اللغة الرئيسية للتدقيق أو مجموعة اللغات
- فصل خطة المواد/الواجهة، إذا لزم الأمر
- إطار اختيار المشاركين

لا يجب أن يتم عمل الخطة والواجهة كطبقات منفصلة لكل دولة. يمكن إدارتها من خلال:

- الطبقة الفرعية،
- الحصة،
- التعيين العشوائي

الطرق.

9. طبقات الدول الكبيرة والصغيرة

الحصة الخام لأي دولة:

```
q_c= nN_c/ N
```

تكون بهذا الشكل. إذا كانت الحصة الخام كبيرة بما فيه الكفاية، يمكن الاحتفاظ بالدولة مباشرة كطبقة منفصلة. يمكن أن يُطلق على هذا:

طبقة الدولة المُمثلة ذاتيًا

تحويل الدول التي لديها حصص أولية صغيرة جدًا بشكل فردي إلى طبقات إلزامية منفصلة قد يقلل من كفاءة السحب. يمكن تصنيف هذه الدول بشكل مناسب حسب:

- إقليميًا،
- عائلة اللغة،
- نظام الوصول إلى المنتجات،
- القرب الجغرافي

يمكن تضمينها في مجموعات الدول الصغيرة.

10. مجموعة الدول الصغيرة

مجموعة الدول الصغيرة:

```
ليتم تمثيلها كـ B_r= {c_1, c_2, …, c_k}
```

إجمالي السكان المؤهلين في المجموعة:

```
N_{B_r}= Σ_{c∈B_r} N_c
```

الحصة الأساسية للعينة:

```
q_{B_r}= nN_{B_r}/ N
```

يتم حسابه كما يلي. داخل التجمع، يمكن اختيار الدولة باحتمال يتناسب مع حجم السكان المؤهلين. احتمال اختيار الدولة c داخل التجمع:

```
Pr(c | B_r) = N_c/ N_{B_r}
```

قد يكون كذلك. بعد ذلك يتم اختيار مشارك داخل الدولة المختارة. هذه الطريقة:

- يمكن أن تمنع الدول الصغيرة من أن يكون لها احتمال صفر في الاختيار ضمن السكان المستهدفين،
- لا تضمن أن كل دولة سيتم اختيارها في موجة معينة،

تحافظ على تقدير السكان العالمي. عدم اختيار دولة في تلك الموجة:

لم يتم اختيارها في عينة السكان

لا يعني: "البلد ليس جزءًا من السكان المستهدفين."

11. قيود تجميع الدول الصغيرة

لا يمكن وضع الدول الصغيرة عشوائيًا في نفس المجموعة لمجرد صغر حجم سكانها. البلدان داخل المجموعة متشابهة بالقدر الذي يكون ذا أهمية:

- الوصول إلى المنتج،
- واجهة المستخدم
- اللغة أو هيكلية المكان,
- الوصول القانوني,
- المنطقة الجغرافية,
- طريقة جمع البيانات

يجب أن يتم تنفيذها. لا يمكن وضع بلد بدون وصول إلى المنتج في نفس المجموعة الرئيسية مع بلد يتمتع بوصول كامل. لا يمكن إخفاء ظروف الاستخدام المختلفة.

12. مجموعة الدول الصغيرة لا تنتج درجة الدولة

ملاحظة واحدة أو اثنتان مختارتان من المجموعة: يمكن استخدامها في المساهمة السكانية العالمية للمجموعة، وقد تشير إلى حدث في دولة واحدة. ومع ذلك، فهي لا تنتج درجة إحصائية منفصلة لدولة صغيرة. إذا كانت الدرجة المستندة إلى الدولة مرغوبة، بالإضافة إلى ذلك:

- مراقب الدولة،
- عينة محددة للدولة،
- عينة تراكمية متعددة الموجات

قد يكون مطلوبًا.

13. التوزيع متناسب مع السكان الصغيرين

التوزيع التناسبي البحث:

يستند إلى الصيغة:

```
q_h= nW_h
```

المزايا:

- من السهل فهمه لتقدير السكان الرئيسيين
- يعطي الوزن بشكل طبيعي للسكان الكبار،

- يمكن أن يقلل اختلافات الوزن،

يمكن أن يقترب من تصميم ذو وزن ذاتي. القيود:

- يمكن أن يقوم بأخذ عينات من عدد قليل جدًا من الدول الصغيرة واللغات،
- قد يكون غير كافٍ لإنتاج درجات المجموعات الفرعية،
- قد يترك المجموعات الصغيرة عالية المخاطر غير مرئية،

يمكن أن يخلق حصصًا كسرية وقريبة من الصفر. لذلك، يمكن أن يكون التخصيص النسبي البحث افتراضيًا قويًا للوحة السكان الرئيسية. إنه لا يحل محل اللوحات التكاملية.

14. تحويل الكسور إلى أعداد صحيحة

الحصص الخام غالبًا ما تكون كسرية. يجب أن يكون التوزيع الإجمالي بالضبط n. يمكن استخدام طريقة الباقي الأكبر كطريقة افتراضية.

14.1. الخطوة الأولى

يتم تقريب كل حصة خام للأسفل:

```
b_h= ⌊q_h⌋
```

14.2. الخطوة الثانية

عدد الملاحظات غير المخصصة:

```
R= n− Σ_h b_h
```

يتم حسابه.

14.3. الخطوة الثالثة

الجزء الكسري من كل طبقة:

```
f_h= q_h− b_h
```

حيث f_h هو الجزء الكسري من q_h.

14.4. الخطوة الرابعة

يتم إضافة الرصدات المتبقية R واحدًا تلو الآخر إلى الطبقات التي تحتوي على أعلى قيم fh.

14.5. حالة التساوي

إذا كانت الأجزاء الكسرية للطبقتين متساوية:

- حجم السكان،
- ترتيب ثابت محدد مسبقًا،
- بذرة عشوائية مسجلة

يجب استخدام قاعدة مستقلة لكسر التعادل، بغض النظر عن النتيجة. لا يمكن اختيار بلد بناءً على النتائج.

15. التخصيص الصحيح التركيبي

العرض التركيبي

السكان المناسبون في بداية القسم:

البلد	السكان المناسبون	الحصة الخام
A	400,000,000	400.0
B	250,000,000	250.0
C	150,000,000	150.0
D	100,000,000	100.0
E	60,000,000	60.0
F	30,000,000	30.0
G	8,000,000	8.0
H	1,500,000	1.5
I	400,000	0.4
J	100,000	0.1

إذا كانت الدول محمية كطبقات منفصلة، قم بالتقريب لأسفل:

البلد	الحصة المقربة لأسفل
A	400
B	250
C	150
D	100
E	60
F	30
G	8
H	1
I	0
J	0
الإجمالي	999

تبقى ملاحظة واحدة. يتم إضافتها إلى H لأن أعلى جزء هو 0.5 في الدولة H. التوزيع النهائي:

البلد	الحصة النهائية
A	400
B	250
C	150
D	100
E	60
F	30
G	8
H	2
I	0
J	0
الإجمالي	1,000

على الرغم من أن هذا التصميم يحدد **I** و **J** كطبقات حتمية منفصلة، إذا لم يمنح فرصة للاختيار، فقد يشكل مشكلة من حيث استنتاج كامل الكون المستهدف. الحل الأكثر دقة هو مجموعة صغيرة من الدول.

16. التوزيع التركيبي باستخدام مجموعة صغيرة من الدول

الدول H و I و J:

```
B_small= {H, I, J}
```

يجب أن يتم دمجه في المجمع. عدد سكان المجمع:

```
1,500,000+400,000+100,000=2,000,000
```

الحصة في المجمع:

```
1,000× 2,000,000/ 1,000,000,000= 2
```

تصبح. التخصيص الرئيسي:

الطبقة	الحصة
A	400
B	250
C	150
D	100
E	60
F	30
G	8
مجمّع الدولة الصغيرة	2
الإجمالي	1,000

أوزان اختيار الدول داخل المجمع:

البلد	حصة السكان داخل المجمع
H	75%
I	20%
J	5%

يتم اختيار ملاحظتين رئيسيتين للسكان وفقًا لهذه الاحتمالات. في نفس الموجة:

- يمكن أن تأتي كلتا الملاحظتين من H,
- إحداهما من H والأخرى من I,
- نادراً من J

يمكن أن يحدث هذا. التصميم لا يجعل **J** مرئيًا في كل موجة، لكنه لا يقلل من احتمال اختيار **J** إلى الصفر. إذا كان يجب ملاحظة **J**، يتم إضافته بشكل منفصل إلى لوحة مراقبي الدولة. ملاحظة مراقب الدولة لا تمنح **J** وزنًا قدره واحد في 1,000 في التوزيع العالمي الرئيسي؛ يحتفظ **J** بوزنه السكاني الحقيقي في اللوحة الرئيسية.

17. الحد الأدنى لتعيين الخلايا

قد لا يكون الحصة النسبية البحتة كافية لإجراء تحليل تشخيصي لبعض الفئات الفرعية. على سبيل المثال، الحصة الخام لخلايا اللغة قد تكون 8. لا يمكن إنتاج درجة لغة موثوقة مع ثمانية ملاحظات. في هذه الحالة، هناك ثلاثة خيارات:

- عدم نشر درجة اللغة
- دمج الخلية مع المجموعات المماثلة
- زيادة أخذ العينات من الخلية

يجب تحديد الحد الأدنى لتخصيص الخلية قبل النتائج.

18. نطاقات الإبلاغ عن المرشحين

نطاقات الإبلاغ عن المرشحين — نطاقات الإبلاغ عن المرشحين

المعايير أدناه ليست قرارات علمية نهائية. يجب معايرتها من خلال تجربة تجريبية ومراجعة خارجية.

1–29 مشاهدات صالحة

الحدث الفردي ومعدل الاكتشاف حسب البلد أو اللغة غير منشور

للمراقبة فقط

30–99 مشاهدات صالحة

تقدير على مستوى استكشافي عدم اليقين واسع لا يُقدَّم حكم قطعي على المطابقة

استكشافي

100–199 مشاهدات صالحة

تقدير مجموعة فرعية أولى يجب أن يكون عدم اليقين مرئيًا

مؤقت

200 وما فوق الملاحظات الصالحة

مرشح تقدير مجموعة فرعية أقوى يتم تقييم تأثير التصميم والأوزان مرة أخرى

مرشح التقدير القياسي

هذه النطاقات:

- معدل الخطأ،
- الدقة المتوقعة،
- تغير الوزن،
- تأثير التصميم

لا يمكن اعتبارها تلقائيًا كافية دون أخذها في الاعتبار.

19. أخذ عينات زائد

قد يكون لدى بلد أو لغة أو مجموعة مستخدمين عدد من الملاحظات أكثر من حصتها من السكان. يُسمى هذا بأخذ العينات الزائدة. أغراضه:

- لزيادة حساسية المجموعة الفرعية،
- لجعل لغة منخفضة الموارد مرئية،
- لفحص مجموعة صغيرة عالية المخاطر،
- لتوسيع تغطية مراقب الدولة،
- لتمثيل مستخدمي إمكانية الوصول بشكل كافٍ

ممكن. الإفراط في العينات ليس خطأ. ما هو خطأ:

```
إنه لاستخدام نسبة العينة من المجموعة التي تم أخذ عينات مفرطة منها كما لو كانت هي نسبة السكان الحقيقية.
```

20. حصة السكان وحصة العينة

حصة السكان للطبقة h:

```
W_h= N_h/ N
```

مشاركة العينة:

```
S_h= n_h/ n
```

في النموذج. في التوزيع النسبي البحت:

```
S_h≈ W_h
```

حسنًا. في الغ_sampling المفرط:

```
S_h> W_h
```

هذا ممكن. في عينة صغيرة:

```
S_h< W_h
```

هذا ممكن. يجب أن يحتفظ تقدير السكان العالمي بقيم Wh بدلاً من Sh.

21. الحمولة الأساسية للتصميم

في العينات الطبقية البسيطة، يمكن اعتبار الوزن الأساسي لكل ملاحظة داخل الطبقة h على النحو التالي:

```
d_i= N_h/ n_h
```

يمكن كتابة وزن الطبقة الموحد على النحو التالي:

```
a_h= W_h/ S_h
```

يمكن أن يكون وزن مستوى الملاحظة الموحد:

```
w_i= W_h/ n_h
```

يتم التأكد من أن المجموع يساوي واحد:

```
Σ_i w_i= 1
```

يمكن حساب تقدير الانتقال الثنائي المرجح على النحو التالي:

```
θ̂= Σ_i w_iY_i
```

هنا:

```
Yi=1: استجابة تلبي معيار الانتقال
Yi=0: استجابة لا تلبي المعيار
```

هذا ممكن. لا يجب على النظام النهائي NOMOS استخدام النتائج الثنائية فقط. يمكن تطبيق نفس منطق الوزن أيضًا على درجات الاستجابة متعددة الأبعاد.

22. مثال على أخذ عينات زائدة

مثال اصطناعي

في السكان المناسبين للدولة ب:

- اللغة X: 80 بالمئة
- اللغة Y: 20 بالمئة

دع الأمر يكون كذلك. في لوحة السكان، هناك 250 ملاحظات للدولة ب. التوزيع البحث:

- اللغة X: 200
- اللغة Y: 50

دع 50 ملاحظة إضافية للعدالة اللغوية تُؤخذ للغة Y. الملاحظات المجموعة:

- اللغة X: 200
- اللغة Y: 100

حصة اللغة Y في العينة الخام:

```
100/300 = 33.3%
```

لقد كان كذلك. الحصة الفعلية للسكان هي 20%. في حساب السكان العالمي، لا يمكن للغة Y أن تأخذ وزن 33.3%. الأوزان المعيارية ضمن الطبقة:

- إجمالي الوزن للغة X: 80%
- إجمالي الوزن للغة Y: 20%

يتم الحفاظ عليه. كل من الملاحظات 100 في لغة Y تتلقى وزنًا فرديًا أقل من ملاحظة واحدة في لغة X. بالمقابل، تُستخدم جميع الملاحظات 100 في تقرير التشخيص المنفصل للغة Y. هذه الطريقة: تزيد من وضوح المجموعة الفرعية، ولا تشوه تقدير السكان العالمي.

23. توزيع المستخدمين متعددي اللغات

قد يكون الشخص مناسبًا لأكثر من خلية لغة واحدة. واحدة من الأساليب المفضلة للوحة السكان الرئيسية هي تعيين لغة تدقيق رئيسية محددة مسبقًا للمستخدم. يمكن تحديد لغة التدقيق الرئيسية وفقًا لهذه الحقول:

- لغة الاستخدام اليومي للذكاء الاصطناعي
- اللغة الأم أو لغة المنزل
- السوق والمنطقة
- الكفاءة في استخدام المُدخل بشكل طبيعي
- توازن اللوحة

إذا كان سيتم اختبار المستخدم أيضًا بلغات أخرى، فتؤخذ هذه الملاحظات في الاعتبار:

- لوحة متعددة اللغات مزدوجة،

- لوحة عدالة اللغة

ويتم تصنيفها وفقًا لذلك. يُحظر أن يحصل الشخص نفسه على وزن سكاني كامل في جميع اللغات.

24. تخصيص اللغة داخل البلد

لنفرض أن السكان المناسبين لمجموعات اللغة l في البلد c هم: N_{c,l}. إذا كانت حصة البلد هي: n_c، فإن الحصة الأولية للغة:

```
q_{c,l}= n_c× N_{c,l}/ Σ_l N_{c,l}
```

يمكن حسابها على هذا النحو. ومع ذلك، إذا كانت مجموعات اللغة تتداخل:

```
Σ_l N_{c,l}> N_c
```

قد يكون ذلك ممكنًا. في هذه الحالة، استخدام الحصة المباشرة يؤدي إلى العد المزدوج. خيارات الحل:

- تصنيف اللغة الأساسي للتدقيق،
- تقسيم الوزن البشري،
- نموذج مشترك متعدد اللغات,
- لوحة اللغة المطابقة المنفصلة

قد يتم استخدامها. يجب تسجيل الطريقة المستخدمة.

25. تخصيص الخطة والواجهة ضمن السكان

قد يتصرف المستخدمون المجانيون والمدفوعون أو واجهات الويب والجوال بشكل مختلف. ومع ذلك، فإن تقسيم كل خلية بلد–لغة بشكل منفصل إلى طبقات الخطة والواجهة يمكن أن يؤدي إلى تجزئة مفرطة للعينة. يمكن استخدام ثلاث طرق.

25.1. التخصيص الجزئي بما يتناسب مع السكان

إذا كانت توزيع المستخدمين الفعلي معروفًا، يتم التخصيص وفقًا لحصص الخطة والواجهة.

25.2. التخصيص التجريبي المتوازن

يتم إجراء أخذ عينات متساوية أو أكبر لمقارنة الفرق بين الخطة أو الواجهة. في التقدير العالمي، يتم إعادة وزنه إلى حصص الاستخدام الفعلية.

25.3. درجات واجهة المنتج المنفصلة

يتم الإبلاغ عن المنتجات المجانية على الويب، والمدفوعة على الويب، ومنتجات الهاتف المحمول كأجسام تدقيق منفصلة. يمكن أيضًا إنتاج نتيجة مجمعة. يجب تحديد النهج الذي سيتم استخدامه قبل القياس.

26. رقم الدعوة والهدف من الملاحظة الصالحة

إذا كان الهدف من الملاحظة الصالحة لفئة معينة هو nh، فإن دعوة nh شخصًا بالضبط عادةً لا تكون كافية. لأن:

- بعض الأشخاص لا يشاركون،
- وبعضهم لا يكمل المهمة,
- بعض السجلات تظل خارج البروتوكول،

بعض حزم الأدلة تصبح غير صالحة. دع معدل الإنجاز المتوقع للفئة يكون rh ومعدل الصلاحية المتوقع يكون vh. الهدف من الدعوة هو تقريبًا:

```
m_h= ⌈n_h/ (r_h v_h)⌉
```

يمكن حسابه كما يلي. مثال:

- هدف الملاحظة الصالحة: 100
- معدل الإنجاز المتوقع: 80 في المئة
- معدل الصلاحية المتوقع: 90 في المئة

```
m_h= ⌈100/ (0.80× 0.90)⌉ = 139
```

يمكن دعوة الأشخاص. هذا ليس ضمانًا. مع تقدم المجال، يمكن استخدام نظام النسخ الاحتياطي المحدد مسبقًا.

27. حد التضخم في الدعوات

دعوة المزيد من الأشخاص وحدها لا تحل مشكلة تحيز العينة. قد يكون المستجيبون الأوائل مختلفين نظاميًا عن المستجيبين المتأخرين. أخذ أول 1,000 من المكتملين فقط:

- المنطقة الزمنية،
- اهتمام التكنولوجيا،
- جودة الاتصال،
- جدول العمل

قد يؤدي إلى تحيز من حيث. ترتيب الدعوة والاختيار:

- يجب أن يكون عشوائيًا،
- يجب تسجيله ضمن الطبقة،

يجب إدارته طوال نافذة المجال.

28. نظام المشارك البديل

يجب إنشاء قائمة نسخ احتياطية مرتبة لكل شريحة قبل رؤية النتائج. يمكن تنشيط مشارك احتياطي فقط للأسباب التالية:

- عدم الاستجابة للدعوة
- عدم إتمام المهمة
- مشكلة في الملاءمة التقنية أو الأخلاقية
- انتهاك بروتوكول الطلب
- عدم صلاحية حزمة الأدلة
- إعادة تسجيل نفس الشخص

لا يمكن للمستخدم الاحتياطي التدخل للأسباب التالية:

```
رد الذكاء الاصطناعي للمستخدم السابق أسفر عن ناتج غير صحيح أو سلبي أو غير مرغوب فيه.
```

الاستجابة غير الصحيحة للذكاء الاصطناعي هي النتيجة النظامية الصحيحة. لا يمكن أن تكون سببًا لتغيير المشارك.

29. العَمَى في النتيجة عند اتخاذ قرار الصلاحية

الشخص الذي يقرر ما إذا كانت الملاحظة تتوافق مع البروتوكول قدر الإمكان:
- يجب أن يقيم الاستجابة دون معرفة ما إذا كانت صحيحة أم خاطئة،
- أو ما إذا كانت إيجابية أم سلبية للعلامة التجارية.

أولاً:
- القبض على الصلاحية والبروتوكول
- ثم يجب تنفيذ تقييم الاستجابة الموضوعية

وإلا، يمكن إعلان الاستجابات السلبية على أنها: "غير صحيحة" بسهولة أكبر.

30. طلب احتياطي داخل الطبقة

يجب أن يكون المشارك الاحتياطي من نفس المصدر قدر الإمكان:
- البلد،
- اللغة،
- واجهة المنتج
- الخطة،
- حالة اللوحة

الطبقة. إذا لم يتبق أي احتياطي في نفس الطبقة، يتم تطبيق التصعيد المحدد مسبقًا:
- اختيار إضافي عشوائي داخل نفس الطبقة
- مجموعة الطبقة العليا المحددة مسبقًا
- ترك التخصيص غير مكتمل وتحذير النطاق
- نسخة تحليل جديدة

بعد رؤية النتيجة، لا يمكن جلب مستخدم من دولة أخرى وسد الفجوة بصمت.

31. ماذا يحدث إذا كانت الطبقة قصيرة جدًا؟

قد ينخفض تحت هدف طبقة. مثال:
- الهدف: 100
- ملاحظة صالحة: 72

الخيارات الصحيحة:
- توسيع نافذة الحقل إلى مدى محدد مسبقًا
- استخدام قائمة النسخ الاحتياطي
- دمج الخلية مع طبقة عليا مشابهة ومحددة مسبقًا
- الإبلاغ عن النتيجة مع درجة غير يقين واسعة
- عدم نشر نتيجة الخلية
- إعادة حساب الوزن العالمي باستخدام طريقة مناسبة

الخيار الخاطئ: إزالة الطبقة منخفضة الدرجات بالكامل من التحليل.

32. تغيير الحصة بعد النتيجة

التغييرات التالية تعتبر تلاعبات نموذجية مرتكزة على النتائج:

- تقليل الحصة المستهدفة للبلد منخفض النقاط
- زيادة حصة اللغة عالية النقاط
- إلغاء صلاحية المستخدمين الذين يقدمون ردودًا سلبية
- إعادة تعريف مجموعة الدول الصغيرة بناءً على النتائج
- استبعاد خطة تحتوي على أخطاء حرجة
- إضافة مستخدمين جدد حتى الوصول إلى المتوسط المرغوب

كل تغيير في حصة المواد:

- نسخة التخصيص الجديدة،
- التبرير،
- تاريخ،
- صاحب القرار
- الحفاظ على النتيجة السابقة

يجب القيام به مع.

33. أخذ العينات متعدد المراحل

يمكن استخدام أخذ العينات متعدد المراحل عندما لا يكون الوصول المباشر إلى جميع الأفراد ممكنًا. مثال:

- اختيار منطقة أو مجموعة دول
- اختيار الدولة
- اختيار لوحة أو شريك بحث محلي
- اختيار المستخدم
- تعيين المنتج أو المهمة الخاصة بالذكاء الاصطناعي

إجمالي احتمالية اختيار مستخدم فردي هو ناتج الاحتمالات الشرطية في كل مرحلة:

```
π_i= π_r× π_{c|r}× π_{f|c}× π_{i|f}
```

هنا:

- r: المنطقة أو مجموعة دولة صغيرة
- c: الدولة
- f: إطار أخذ العينات أو اللوحة
- i: الشخص

وزن التصميم الأساسي:

```
d_i= 1/π_i
```

إذا لم يكن من الممكن تسجيل احتمالات الاختيار، فلا يمكن تقديم التصميم كعينة قائمة بالكامل على الاحتمالات. [K04; K07]

34. التوزيع في اللوحات الطوعية وغير المعتمدة على الاحتمالات

قد لا يكون الحصول على لوحة عالمية قائمة على الاحتمالات دائمًا ممكنًا. يمكن استخدام لوحة طوعية أو مهنية على الإنترنت. في هذه الحالة، لا يزال من الممكن تحديد حصص الدولة واللغة وفقًا للسكان المؤهلين للمنتج. ومع ذلك:

- قد تكون الاحتمالية الفعلية لاختيار المشارك غير معروفة.
- قد يحدث اختيار ذاتي,

يمكن للترجيح إجراء تعديلات بناءً فقط على المتغيرات المرصودة. الحالة الصحيحة:

لوحة مستخدم غير احتمالية مُقسمة ومعايرة للسكان

قد تكون. البيان التالي لا يمكن استخدامه: "1,000 أشخاص مختارون بشكل عشوائي بالكامل من سكان العالم."

35. هيكلية الترجيح

قد يتكون وزن الملاحظة النهائي من عدة مكونات:

```
w_i= d_i× a_i^{NR}× g_i^{CAL}
```

هنا:

- di: وزن التصميم الأساسي
- aiNR: تعديل عدم الاستجابة
- giCAL: معامل المعايرة

35.1. وزن التصميم

إنه مقلوب احتمال اختيار المشارك في تصميم العينة.

35.2. تعديل عدم الاستجابة

للأشخاص المختارين ذوي الخصائص المماثلة:

- كم منهم أجاب،
- كم منهم أنتج ملاحظة صالحة

يمكن أخذها في الاعتبار. لفئة استجابة العينة r:

```
a_r^{NR}= (Σ_{j∈S_r} d_j)/ (Σ_{i∈R_r} d_i)
```

هنا:

- Sr: العينة المؤهلة المختارة
- Rr: الذين استجابوا أو أنتجوا ملاحظة صالحة

قد يكون. تعديل عدم الاستجابة: لا يحل تمامًا التحيز غير المرصود، ويجب شرحه مع افتراضاته.

35.3. وزن المعايرة

العينة، إلى إجماليات السكان المعروفة:

- البلد،
- اللغة،
- العمر،

- الخطة،
- واجهة المستخدم.

يمكن تطبيق المعايرة عبر هذه الأبعاد، ولكن فقط حيث توجد إجماليات خارجية موثوقة.

36. تطبيع الوزن

الأوزان:

- إلى إجمالي السكان،
- إلى حجم العينة الكلي،
- إلى الفرد

يمكن تطبيعها. يجب شرح التطبيع المستخدم في صيغة النتيجة. الأوزان المطابقة بشكل فردي:

```
w̃_i= w_i/ Σ_j w_j
```

تم الحصول عليها. التقدير المرجح:

```
θˆ= Σ_i w̃_iY_i
```

يتم حسابها كما يلي.

37. تقليم الأوزان

قد يتلقى بعض المشاركين أوزانًا عالية جدًا لأنهم يمثلون مجموعات صغيرة ونقصت عيناتها. هذا الوضع:

- يمكن أن يزيد من التباين المقدر،
- أثر مشاهد واحدة،
- وتقلب الدرجات

يمكن أن يزيد. يمكن تقليص الأوزان عند حد أعلى. ومع ذلك، فإن التقليص: قد يخلق تحيزًا في تمثيل السكان، قد يجعل المجموعات الصغيرة غير مرئية مرة أخرى. لذلك، يجب أن يكون تقليص الأوزان:

- مُعرفًا قبل النتائج،
- معروضًا مع النتائج غير المُقلصة،
- ويحمل تحليل الحساسية.

38. حجم العينة الفعّال

على الرغم من أن العدد الخام للملاحظات هو n، إذا كانت الأوزان تختلف بشكل كبير، قد تكون كمية المعلومات الفعلية أقل. حجم العينة الفعّال وفقًا لـ Kish:

```
n_eff= (Σ_i w_i)^2/ Σ_i w_i^2
```
[K08]

يمكن عرضه كما يلي. مع الأوزان المتساوية:

```
neff≈n
```

يتم الحصول عليه. إذا كانت الأوزان غير متوازنة جدًا: neff<n. دراسة GEO-1000:

- 1,000 من الملاحظات الصالحة،
- قد تحمل 620 حجم العينة الفعّال

للجمهور فقط: قول "تم قياس 1,000 شخص" قد يظهر قويًا بسبب الدقة. في التقرير الرئيسي، على الأقل: .

- عدد الملاحظات الصالحة الخام،
- حجم العينة الفعّال،
- تباين الأوزان

يجب تحديده.

39. تأثير تصميم الوزن

التأثير التقريبي للتصميم بسبب الأوزان:

```
DEFF_w≈ 1+ CV(w)^2
```

يمكن إظهار ذلك. هنا:

```
CV(w) = sd(w) / mean(w)
```

هذا هو معامل التباين للأوزان ويمثل فقط تقريبًا تأثيرها. قد يختلف التأثير الكلي للتصميم أيضًا مع:

- التجمع,
- التصنيف,
- القياس المتكرر,
- المستخدمون المزدوجون

يجب تقدير هذه التأثيرات بشكل منفصل.

40. التجمع

نفس الشيء:

- من لوحة البحث,
- من البلد,
- من مجتمع اللغة،
- من الجهاز أو المؤسسة،
- من المستخدم

قد تتشابه الملاحظات بعضها مع بعض. وفي هذه الحالة، لا تمثل 1,000 ملاحظة 1,000 وحدة معلومات مستقلة تماماً. التجميع العنقودي:

- في تصميم العينة،
- في حساب التباين،
- في حجم العينة الفعّال

يجب أخذها في الاعتبار.

41. اختبار أكثر من منتج ذكاء اصطناعي واحد بواسطة نفس المستخدم

إجراء اختبار مستخدم لعدة منتجات ذكاء اصطناعي يسمح بقياس الفرق بين النماذج على نفس الشخص. الميزة: يمكن التحكم في خصائص المستخدم، ويمكن تعزيز المقارنة المزدوجة. القيد:

- ملاحظات المنتج غير مستقلة،
- ترتيب المهام يمكن أن يؤثر على استجابات أخرى،

- قد ينقل المستخدم الإجابة التي شاهدها في النظام الأول إلى النظام التالي،

قد يختلف المستخدمون الذين يمكنهم الوصول إلى جميع المنتجات عن المجتمع السكاني. لذلك، يجب على المنتجات:

- تُخصص بترتيب عشوائي،
- يجب منع مشاركة الإجابات السابقة،
- يجب تسجيل التصميم المزدوج.

يجب أن يحمي التحليل الاعتماد. يجب أن يكون لكل منتج ذكاء اصطناعي مجددًا هدفه الخاص من ملاحظات المنتجات الصالحة .1,000 إذا اختبر نفس البشر 1,000 عشرة منتجات:

- 1,000 البشر الفريدون،
- ملاحظات المنتج 10,000

يمكن تكوينها. هذان الرقمان ليسا متساويين.

42. تصميم الكتل الناقصة المتوازن

ليس كل المستخدمين بحاجة للوصول إلى جميع منتجات الذكاء الاصطناعي. على سبيل المثال، إذا كان هناك عشرة منتجات ذكاء اصطناعي، يمكن تخصيص ثلاثة منتجات لكل مشارك. إذا تم إجراء التعيينات بطريقة متوازنة:

- كل منتج 1,000 ملاحظات،
- كل زوج من المنتجات مستخدمون مشتركين كافون
- مشاركة أوسع من قبل السكان

يمكن تحقيقها. كمرشح لهذا النهج:

تصميم الكتل غير الكامل المتوازن

يمكن استخدامه. مميزاته:

- يقلل من عبء المشاركين,
- يخفض خطر اختيار المستخدمين الثقيلين فقط للذكاء الاصطناعي,

يوفر مقارنة متطابقة عبر النماذج. سيتم تفصيل التصميم الدقيق في الأقسام 8 و .9

43. نطاق الوصول الطبيعي التوزيع

يحصل كل منتج ذكاء اصطناعي على تخصيص منفصل وفقًا لسكانه الخاصين في نطاق الوصول الطبيعي. وبالتالي، نفس الدولة:

- 200 لمنتج واحد,
- 0 لمنتج آخر،
- 50 لمنتج ثالث

يمكن ملاحظته. هذا الاختلاف ليس ظلماً. إنه أن عوالم الوصول الحقيقية للمنتجات مختلفة. بما أن درجات نطاق الوصول الطبيعي تنتمي إلى جماعات سكانية مختلفة، فلا ينبغي استخدامها بمفردها للترتيب المباشر.

44. قاعدة المقارنة المشتركة التوزيع

يتم إنشاء عالم وصول مشترك للمقارنة بين المنتجات. الجماعة السكانية المشتركة:

```
U_CS= ⋂_{a∈A} U_a
```

يمكن النظر إليها بهذه الطريقة. يتم قياس كل منتج بنفس أوزان السكان قاعدة المقارنة المشتركة. هذه الطريقة:

- توفر مقارنة عادلة بين المنتجات،
- نفس المستوى في الدولة واللغة

يضمن ذلك. ومع ذلك، فإنه لا يظهر الفارق الطبيعي في الوصول العالمي للمنتجات. لذلك، يجب عرض نتائج نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة معًا.

45. إدارة العينة عبر الموجات

قد يكون هناك هدفان مختلفان عبر موجات القياس المتعددة.

45.1. نتيجة المقطع العرضي الحالي

في كل موجة، يتم استخدام إطار السكان المؤهلين للمنتج الحالي. هذا المقياس يجيب على السؤال: "ماذا يحدث الآن في الكون الفعلي للوصول العالمي؟"

45.2. نتيجة اتجاه الإطار الثابت

تظل أوزان السكان من الموجة الأولية محفوظة. يجيب هذا المؤشر على السؤال: "كيف تغير التمثيل في نفس السكان المرجعيين مع مرور الوقت؟"

45.3. فصل بين نتيجتين

عندما يمتد المنتج إلى دول جديدة، قد ينخفض المؤشر الحالي. هذا الانخفاض:

- يرجع إلى تدهور المنتج،
- نتيجة لإدراج سكان جدد وصعوبتهم

قد يكون السبب في ذلك. يمكن أن يظهر الاتجاه ضمن إطار ثابت تحسناً في نفس الوقت. لذلك:

- لقطة للإطار الحالي
- اتجاه الإطار الثابت

إنها نتائج منفصلة.

46. تجديد وتحول اللوحة

بين الموجات:

- نفس المستخدمين،
- مستخدمين جدد تمامًا،
- لوحة متجددة جزئيًا

قابلة للاستخدام.

46.1. لوحة القسم الكاملة

يتم اختيار مستخدمين جدد في كل موجة. يمكن أن تكون قوية لتقدير السكان. إنها لا تقيس التغيير الفردي.

46.2. لوحة طولية

يتم قياس نفس المستخدمين مرة أخرى. يُظهر تغييرات المستخدم الفردي والمنتج. قد تخلق اللوحة تأثيرات التآكل والتعلم.

46.3. لوحة دوّارة

يتم الاحتفاظ بجزء من المشاركين، في حين يتم تجديد جزء آخر. يوفر تمثيلًا لكل من الاتجاهات والسكان الجدد. يجب أن يكون هيكل اللوحة المستخدم مرئيًا على بطاقة النتائج.

47. تثبيت العينة

يجب تثبيت السجلات التالية قبل بدء كل موجة:

- منتج الذكاء الاصطناعي
- واجهة المستخدم
- نسخة السكان المستهدفين
- نطاق الوصول الطبيعي الدول
- قاعدة المقارنة المشتركة الدول
- الطبقات
- مجموعات الدول الصغيرة
- الحصص الأولية
- التخصيصات الصحيحة
- تخصيصات اللغة
- خطة الفرز الزائد
- أهداف الدعوة
- قوائم الاحتياطي
- طريقة الاختيار
- معادلة الوزن
- قاعدة تعطيل النتائج
- تصعيد النقص
- خطة التحليل
- طريقة فك الترميز
- البذرة العشوائية، إذا تم استخدامها
- موافقة الإنسان المسؤول

لهذا الملف:

تثبيت تخصيص العينة **NOMOS**

يمكن إعطاؤه اسماً.

48. التغييرات مسموح بها بعد تثبيت العينة

قد تتطلب الأحداث التالية إصدار نسخة جديدة:

- إنهاء الوصول إلى المنتج في البلد
- مشكلة خطيرة في سلامة الميدان
- خطأ مادي في بيانات السكان

- مقدم اللوحة غير قادر على التسليم إلى دولة معينة
- تغيير في متطلبات أخلاقيات البحث
- خطأ حرج في أداة القياس

تغيير:

- يجب أن يكون مبررًا،
- بتاريخ،
- معتمدًا،
- يتم تنفيذه مع الحفاظ على النسخة القديمة.

قوة أو ضعف النتائج ليست سببًا للتغيير.

49. حالات جودة العينة

يمكن أن تحمل موجة العينة واحدة من الحالات التالية:

SQ-0 — غير مصممة

لا يوجد تصميم عينة موثق.

SQ-1 — مراقبة بالراحة

هناك مراقبة بالراحة أو طوعية. لا يمكن تعميم النتائج على السكان.

SQ-2 — التحكم بالحصة

هناك حصص للبلد واللغة. احتمالات الاختيار غير معروفة بالكامل.

SQ-3 — لوحة مرجحة

هناك لوحة مُقسّمة ومعايرة للسكان. لا تزال هناك حدود للتغطية والاختيار الذاتي.

SQ-4 — بناء قائم على الاحتمالات أو هجين

توجد احتمالات اختيار معروفة أو تصميم هجين قوي.

SQ-5 — مكرر ومعاير

تم تكرار التصميم في موجات وتطبيقات خارجية مختلفة؛ وتمت معايرة سلوك الانحياز والتباين. قد تؤثر هذه الحالات على فئة الثقة للنتيجة العامة.

50. تخصيص نهائي اصطناعي شامل GEO-1000

عرض المنهجية الاصطناعية / الدول والسكان واللغات ومنتج الذكاء الاصطناعي والحسابات التالية كلها خيالية بالكامل. بالنسبة لمنتج الذكاء الاصطناعي Synthetic Orion، فليكن السكان المناسبون لـ نطاق الوصول الطبيعي هم 1 مليار شخص. هدف لوحة السكان الرئيسية:

```
n=1,000
```

50.1. تخصيص الدول والدول الصغيرة

الطبقة	السكان المناسبون	حصة السكان	الحصة الرئيسية

الطبقة	السكان المناسبون	حصة السكان	الحصة الرئيسية
A	400,000,000	40%	400
B	250,000,000	25%	250
C	150,000,000	15%	150
D	100,000,000	10%	100
E	60,000,000	6%	60
F	30,000,000	3%	30
G	8,000,000	0.8%	8
مجمّع الدولة الصغيرة	2,000,000	0.2%	2
الإجمالي	1,000,000,000	100%	1,000

50.2. تخصيص اللغة في الدولة ب

السكان المؤهلون في الدولة ب:

- اللغة X: 80 بالمئة
- اللغة Y: 20 بالمئة

تخصيص عينة السكان:

- اللغة X: 200
- اللغة Y: 50

تخصيص إضافي للعدالة اللغوية: اللغة ي: 50 هدف الحقل في الدولة ب:

- اللغة X: 200
- اللغة Y: 100
- الإجمالي: 300

ومع ذلك، أوزان سكان عينة السكان العالمية:

- اللغة X: 80% من الدولة ب
- اللغة ي: 20% من الدولة ب

تبقى كما هي.

50.3. ملحق مراقب الدولة

هناك ملاحظتان لاحتمالية H و I و J في مجموعة البلدان الصغيرة في لوحة السكان. كما تتضمن لوحة مراقب الدولة:

H: 1

I: 1

J: 1

يمكنها استهداف الملاحظات. هذه الملاحظات الثلاثة:

- تزيد من التغطية الجغرافية،
- تُظهر أحداث كل دولة على حدة،

لا تُنشئ درجة للبلد.

50.4. ملحق إمكانية الوصول

إذا تبين أن اللوحة الرئيسية لا تغطي مستخدمي الوصول بشكل كافٍ، يمكن أخذ ملاحظات إضافية .40 هؤلاء المستخدمون: يتم معايرتهم وفقًا لمجموعات السكان ذات الصلة، ويقدمون نتيجة لوحة الوصول بشكل منفصل.

50.5. إجمالي حجم المجال

العينة	الهدف الحالي
لوحة السكان	1,000
ملحق العدالة اللغوية	50
ملحق رصد الدولة	3
ملحق الوصول	40
إجمالي الهدف الميداني الصحيح	1,093

الاسم GEO-1000 لا يزال صالحًا. لأن:

> *1,000* هو الهدف الصحيح للملاحظة للوحة السكان الرئيسية.

يمكن إضافة لوحات العدالة التكميلية والتغطية إلى هذا.

50.6. حجم الدعوة

في لوحة السكان في المتوسط:

- الإكمال: 82 في المئة
- الصلاحية: 92 في المئة

ليكن متوقعاً. الدعوة المطلوبة:

```
⌈1,000/ (0.82× 0.92)⌉ = 1,326
```

قد يكون. مع اللوحات التكميلية، قد تتجاوز إجمالي الدعوات .1,500 لذلك:

> *GEO-1000* ليست فقط طريقة لإرسال الدعوات إلى *1,000* شخصًا.

50.7. نتيجة الميدان

في لوحة السكان:

- تم دعوة 1,326 أشخاص،
- تم إكمال مهام 1,080،
- تم العثور على سجلات 1,012 صالحة من حيث البروتوكول،
- تم تضمين أول 1,000 في التحليل الرئيسي وفقًا لترتيب عشوائي/احتياطي محدد مسبقًا

حسنًا. السجلات الصالحة المتبقية لـ 12:

- التحقق الاحتياطي،
- تحليل الحساسية،
- إعادة التحقق

يمكن تخزينه. لا يمكن تحديد أي 1,000 سيتم اختياره من خلال النظر في النتائج.

51. خوارزمية التوزيع المعيارية

التسلسل الافتراضي للعمليات للوحة الرئيسية GEO-1000 هو كما يلي:

الخطوة 1— تثبيت المنتج ومعرّف الموجة

تحديد أي منتج AI، وما الواجهة، والتاريخ الذي تم قياسه.

الخطوة 2— تثبيت نسخة الفئة المستهدفة

يتم استخدام إطار السكان المؤهلين للمنتج من القسم 5.

الخطوة 3— تحديد الهدف الرئيسي للملاحظة الحالية

الهدف المرشح الافتراضي:

```
n=1,000
```

الخطوة 4— تحديد طبقات المواد

يتم تحديد الدولة واللغة، وإذا لزم الأمر، طبقات الخطة/الواجهة.

الخطوة 5— إعداد تجمعات الدول الصغيرة

يتم تضمين الدول التي ستولد تخصيصات قريبة من الصفر في التجمعات المناسبة.

الخطوة 6— حساب الحصص النسبية الأولية

```
qh=nWh
```

الخطوة 7— تنفيذ التخصيص الصحيح

يتم تطبيق طريقة الباقي الأكبر أو الطريقة المكافئة المعلنة مسبقًا.

الخطوة 8— إضافة خطة أخذ العينات الزائدة

يتم حفظ ملحقات مراقب الدولة، وعدالة اللغة، وإمكانية الوصول بشكل منفصل.

الخطوة 9— حساب أعداد الدعوات والاحتياطيات

يتم استخدام معدلات الاكتمال المتوقعة والصلاحية.

الخطوة 10— عشوائية ترتيب الاختيار

القوائم الرئيسية والنسخ الاحتياطية مجمدة.

الخطوة 11— حفظ صيغة الوزن

تم تعريف أوزان التصميم والاستجابة والمعايرة.

الخطوة 12— تأكيد تثبيت العينة

يتم إنشاء سجل النسخة والنزاهة قبل عرض استجابات الذكاء الاصطناعي.

52. الأحكام الإلزامية المعيارية

CH06-N01

يجب إنشاء تخصيص عينة GEO-1000 وإصدار نسخ قبل عرض استجابات الذكاء الاصطناعي.

CH06-N02

يجب تعريف هدف لوحة السكان الرئيسية كحد أدنى 1,000 ملاحظات صالحة لكل منتج ذكاء اصطناعي وموجة القياس الرئيسية؛ إذا تم استخدام هدف مختلف، يجب شرح المبرر.

CH06-N03

يجب التمييز بين الشخص المدعو، والشخص الذي يكمل المهمة، وأرقام الملاحظات الصالحة.

CH06-N04

يجب أن يكون التخصيص الرئيسي بناءً على السكان المؤهلين للمنتج المحددين في القسم 5.

CH06-N05

يجب حساب حصص الدول أو اللغات بناءً على نسبة السكان المؤهلين للمنتج المعني، وليس على إجمالي السكان.

CH06-N06

يجب تحديد طريقة تحويل الحصص الكسرية الخام إلى أعداد صحيحة مسبقًا.

CH06-N07

يجب أن يكون حل تعادل التقريب مستقلًا عن النتائج.

CH06-N08

لا ينبغي ترك الدول الصغيرة ذات السكان المستهدفين الإيجابيين كطبقات حتمية منفصلة ذات احتمال صفر؛ ينبغي استخدام طريقة تجميع مناسبة أو اختيار احتمالي.

CH06-N09

يجب تبرير تجمعات البلدان الصغيرة من حيث الوصول إلى المنتج والخصائص المالية للسكان.

CH06-N10

يجب أيضًا تسجيل ملاحظات المراقب الوطني دون تغيير الوزن السكاني الرئيسي.

CH06-N11

لا يمكن تقديم ملاحظة مراقب بلد على أنها درجة للبلد.

CH06-N12

يجب إعادة البلدان أو اللغات أو مجموعات المستخدمين التي تم أخذ عينات مفرطة منها إلى الأوزان الحقيقية للسكان المستهدفين في التقدير العالمي.

CH06-N13

لا يمكن استخدام نسبة العينة مثل نسبة المجتمع.

CH06-N14

لا يمكن تكرار نفس الشخص في عدد السكان البشريين العالمي بسبب تعدد اللغات أو الحسابات أو الجلسات أو المنتجات.

CH06-N15

يجب تصنيف الملاحظات غير الصالحة دون الاعتماد على محتوى النتائج كمبرر لعدم الصلاحية.

CH06-N16

لا يمكن أن تكون الاستجابات غير الصحيحة أو السلبية أو الرفضية أو الاستجابات الحرجة للذكاء الاصطناعي سببًا لتغيير المشارك.

CH06-N17

يجب إنشاء قوائم المشاركين الاحتياطية ضمن الطبقة قبل الاطلاع على النتائج.

CH06-N18

لا يمكن لدولة أو لغة أخرى إغلاق فجوات الطبقة بهدوء.

CH06-N19

تتطلب تغييرات الحصة بعد النتائج أو الطبقة أو مجموعة الدول الصغيرة أو الأوزان إصدار تحليل جديد.

CH06-N20

يجب نشر جميع مكونات الأوزان وطريقة التطبيع.

CH06-N21

يجب توضيح افتراضات تعديل عدم الاستجابة.

CH06-N22

إذا تم استخدام تقليم الأوزان، يجب الاحتفاظ بالنتيجة غير المقلصة وتحليل الحساسية.

CH06-N23

يجب أن يوضح التقرير الرئيسي حجم العينة الفعّال المناسب جنبًا إلى جنب مع العدد الخام للملاحظات.

CH06-N24

يجب أخذ التكتل والتكرار واعتماديات المستخدمين المرتبطين في حساب التباين في الاعتبار.

CH06-N25

لا يمكن عرض التصميم الذي يختبر فيه نفس الشخص عدة منتجات ذكاء اصطناعي كعينة مستخدم مستقلة تمامًا.

CH06-N26

يجب على كل منتج ذكاء اصطناعي أن يحقق بشكل منفصل هدفه الخاص من 1,000 من الملاحظات الصالحة للمنتج.

CH06-N27

يجب أن تحتفظ المخصصات نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة بسجلات عينة منفصلة.

CH06-N28

لا يمكن الجمع بين النتائج الحالية وثابتة الإطار بين الموجات دون تفسير.

CH06-N29

إذا كانت العينة غير كافية لمجموعة فرعية، فيجب أن تكون الحالة الصحيحة هي عينة غير كافية، استكشافية، أو غير مقدرة.

CH06-N30

لا يمكن استخدام حجم العينة الخام كمؤشر دقة نهائي، متجاهلين الأوزان وتأثير التصميم.

CH06-N31

لا يمكن المطالبة بأوزان التصميم في عينة تطوعية أو عينة غير عشوائية كما لو كانت احتمالات الاختيار معروفة.

CH06-N32

لا يمكن تعديل العينة بعد الحدث لاستبعاد المجموعات منخفضة الدرجات أو لتكبير المجموعات مرتفعة الدرجات.

CH06-N33

يجب تصنيف الملاحظات من اللوحة السكانية الرئيسية واللوحات التكميلية بشكل منفصل من حيث الغرض والوزن.

CH06-N34

يجب أن يكون لتثبيت العينة مالك بشري أو مؤسسي مسؤول.

53. أشكال الفشل

CH06-F01 — تعيين الحصة بعد النتيجة

يتم تحديد عدد الدول واللغات بعد عرض ردود الذكاء الاصطناعي.

CH06-F02 — استخدم النسخة الاحتياطية حتى الحصول على أفضل إجابة

يتم استبدال الردود السلبية بمستخدم جديد حتى لو كانت صالحة.

CH06-F03 — عد دعوات 1,000 كملاحظات صالحة 1,000

عدم الاكتمال وعدم الصلاحية مخفيان.

CH06-F04 — إعادة ضبط الدولة إلى الصفر مع التقريب النقي

تُترك احتمالية اختيار الدولة ذات السكان المستهدفين الإيجابيين عند الصفر.

CH06-F05 — إعطاء شخص واحد لكل دولة وتسميته نسبة السكان

يتم إخفاء الاختلال الناجم عن الحصة الدنيا.

CH06-F06 — اعتبار مراقبة الدولة كمراقبة سكانية

تحصل المراقبة الإضافية لدولة صغيرة على الوزن الكامل في اللوحة الرئيسية لـ 1,000 شخصًا.

CH06-F07 — إنشاء مجموعة الدولة الصغيرة بناءً على النتائج

يتم نقل دولة تحقق درجات منخفضة إلى مجموعة أخرى أو استبعادها من التغطية.

CH06-F08 — وضع دول غير متوافقة في نفس المجموعة

يتم دمج الدول التي تختلف في الوصول إلى المنتجات أو اللغة أو الاختصاص القضائي فقط لأنها صغيرة.

CH06-F09 — المبالغة في وزن أخذ عينات مفرطة

يحدد النتيجة العالمية بحصة من العينة أعلى من حصة اللغة أو حصة سكان الدولة.

CH06-F10 — مضاعفة عدد الأفراد متعددي اللغات

يتم احتساب الشخص كوحدة سكانية منفصلة في كل خلية لغة.

CH06-F11 — اعتبار حصة العينة هي حصة السكان

يتم تحويل أرقام الاستجابة الخام مباشرة إلى النتيجة العالمية.

CH06-F12 — ترتيب الدعوات حسب سهولة الترتيب

المستخدمون الذين يردون أولاً وبسرعة يحددون العينة.

CH06-F13 — استخدام الاحتياطيات من الطبقات المستبعدة

الملاحظات المفقودة عن الدولة أو اللغة يتم استكمالها بهدوء بمشاركين من مجموعة أخرى.

CH06-F14 — ربط البطلان بمحتوى الاستجابة

يتم إعلان استجابة الذكاء الاصطناعي غير الصحيحة باطلًا من الناحية التقنية.

CH06-F15— استبدال الرفض بمستخدم جديد

يتم إزالة رفض النظام من المقام على الرغم من أنه النتيجة الفعلية للمستخدم.

CH06-F16— إخفاء صيغة الوزن

النتيجة العامة لا تظهر بالضبط مع أي السكان وأوزان العينة تم إنتاجها.

CH06-F17— تقليم الأوزان لحذف المجموعات الصغيرة

يتم تقليم المجموعات الصغيرة ذات الأوزان العالية لأن النتيجة لا تبدو جيدة.

CH06-F18— إخفاء العينة الفعالة

حتى إذا انخفضت سجلات 1,000 الخام إلى نقاط بيانات 600–500 بسبب الأوزان، يتم تقديمها كـ "دقة شخص 1,000".

CH06-F19— اعتبار التجميع كاستقلالية

تعتبر الملاحظات من نفس اللوحة أو المؤسسة أو المستخدم مستقلة تمامًا.

CH06-F20— اعتبار نفس 1,000 الناس كعشرة آلاف شخص مستقل

يتم وصف عشر ملاحظات لمنتج ذكاء اصطناعي كما لو أن السكان كانوا أكبر بعشر مرات.

CH06-F21— اعتبار قاعدة المقارنة المشتركة كـ نطاق الوصول الطبيعي

يتم عرض عينة المقارنة الشائعة كما لو كانت هي السكان الفعليين لانتشار المنتج على مستوى العالم.

CH06-F22— تصنيف نتائج نطاق الوصول الطبيعي مباشرة

يتم مقارنة درجات المنتج في مجموعات مستهدفة مختلفة كما لو كانت من نفس الكون.

CH06-F23— إخفاء تغيير السكان في الموجات

على الرغم من دخول دول جديدة أو مجموعات مستخدمين جديدة في التقييم، يتم تفسير الاتجاه فقط على أنه تغيير في جودة المنتج.

CH06-F24— إنتاج درجات محددة من عينات غير كافية لمجموعة فرعية

يتم نشر عدد قليل من الملاحظات كدرجات محددة لدولة أو لغة.

CH06-F25— تغيير تثبيت العينة بصمت

يتم تحديث الحصة أو الوزن أو تعريفات المجموعة دون سجل للنسخة.

CH06-F26— مضاعفة معدلات الأهلية بشكل أعمى

تتجاهل حسابات السكان المؤهلين والحصص خصائص السكان التابعين.

CH06-F27— استبدال اللوحة الرئيسية باللوحة التكميلية

يتم استخدام ملاحظات عدالة اللغة أو مراقب البلد لتقليل الهدف السكاني الرئيسي لـ 1,000.

CH06-F28— حذف الخلايا منخفضة الدرجات عندما تكون اللوحة قصيرة

يتم استبعاد العينات غير المكتملة لأنها غير مفيدة من حيث النتائج.

CH06-F29— استخدام الترجيح كلمسة علمية

يتم الادعاء بأن مشكلة التمثيل تُحل فقط بإضافة الوزن دون شرح الاختيار والانحياز في التغطية.

CH06-F30— الاعتقاد بأنه يجب طلب من GEO-1000 أن يسأل بالضبط 1,000 شخصًا

يتم تجاهل الدعوة، الألواح التكميلية، بطلان الصلاحية، التكرار، وهيكل العينة الفعّال.

54. إجراء التدقيق

الخطوة 1— تسجيل غرض العينة

يتم تقسيم تقدير السكان العالمي إلى تشخيص الفئات الفرعية، نطاق البلد، عدالة اللغة، الوصول، وأغراض مقارنة المنتجات.

الخطوة 2— التحقق من إصدار السكان المستهدفين

يتم مراجعة ملف السكان وتاريخ التسجيل في القسم 5.

الخطوة 3— تحديد الهدف الرئيسي للملاحظة الحالية

يتم كتابة الهدف الافتراضي GEO-1000 أو البديل المبرر.

الخطوة 4— إنشاء الطبقات

يتم تحديد الاختلافات في البلد واللغة والمادة واجهة المستخدم.

الخطوة 5— تحديد مجموعات الدول الصغيرة

لكل مجموعة:

- الدول،
- السكان،
- مبرر التشابه،
- طريقة الاختيار الداخلية

يتم تسجيلها.

الخطوة 6— حساب الحصص الخام

```
qh=nWh
```

يتم استخدام الصيغة أو بديل مبرر.

الخطوة 7— تطبيق التخصيص الصحيح

يتم تسجيل قواعد التقريب وكسر التعادل.

الخطوة 8— إضافة العينة التكميلية الزائدة

يتم تعريف أهداف مراقب الدولة، وعدالة اللغة، والوصول بشكل منفصل.

الخطوة 9— حساب حجم الدعوة والحجز

يتم استخدام توقعات الاكتمال والصلاحية.

الخطوة 10— تجميد الاختيار وحجز الطلب

يتم اختيار أو ترتيب المشاركين قبل رؤية النتائج.

الخطوة 11— حساب أوزان التصميم

يتم تسجيل احتمالات الاختيار وأوزان الطبقة.

الخطوة —12 تحديد الإجراء للبقاء صامتًا وعدم الرد

يتم كتابة استجابة تم الإعلان عنها سابقًا على فجوة الطبقة.

الخطوة —13 إنشاء تثبيت العينة

يتم إضافة معرف المستند والإصدار والتاريخ وسجل التكامل.

الخطوة —14 مراجعة تدفق صلاحية ما بعد الميدان

تم التحقق من أن قرارات الصلاحية تُعطى بشكل مستقل عن دقة الردود.

الخطوة —15 توليد الأوزان النهائية

يتم فصل التصميم إلى مكونات الاستجابة والمعايرة.

الخطوة —16 حساب حجم العينة الفعّال

يتم الإبلاغ عن الفرق بين الرقم الخام وحجم المعلومات.

الخطوة —17 مقارنة التخصيص والنتائج الفعلية

يتم عرض أعداد الملاحظات المخطط لها والفعلية على أساس خلية بخليتها.

الخطوة —18 مراجعة التغييرات بعد النتيجة

تم التحقق من تغييرات الحصص والمجموعة والوزن والنطاق.

55. الأدلة المطلوبة

هوية منتج الذكاء الاصطناعي؛ موجة القياس؛ نسخة السكان المستهدفة؛ هدف لوحة السكان؛ هدف مراقب الدولة؛ هدف عدالة اللغة؛ هدف لوحة الوصول؛ قائمة الطبقات؛ سكان الطبقات؛ أوزان السكان؛ الحصص الأولية؛ طريقة التقريب؛ سجلات التخصيص الكسري والمتبقي؛ قاعدة كسر التعادل؛ مجموعات الدول الصغيرة؛ احتمالات الاختيار ضمن المجموعة؛ تخصيصات الدولة واللغة؛ تخصيصات الخطة والواجهة؛ مبررات أخذ العينات الزائدة؛ أهداف الدعوة؛ افتراضات الإكمال؛ افتراضات الصلاحية؛ قوائم المشاركين الأساسيين واحتياطيين؛ سجل الاختيار أو العشوائية؛ سجلات عدم الاستجابة؛ أسباب المشاهدات غير الصالحة؛ تواريخ قرارات الصلاحية؛ أوزان التصميم؛ تعديلات عدم الاستجابة؛ معاملات المعايرة؛ سياسة تقليم الأوزان؛ أحجام العينات الخام والفعالة؛ التجميع وسجلات المستخدم المطابقة؛ تخصيص نطاق الوصول الطبيعي؛ وتخصيص قاعدة المقارنة المشتركة.

أوزان الإطار الثابتة والحالية مقارنة التخصيص المخطط والفعلي تثبيت العينة سجل التغيير الشخص أو المؤسسة المسؤولة

56. قائمة تدقيق التدقيق

هل الهدف الأساسي للمراقبة صالح ومفتوح؟ هل الهدف 1,000 يدعو إلى الملاحظات أو الصالحة منها؟ هل تم تجميد العينة قبل مشاهدة استجابات الذكاء الاصطناعي؟ أي إطار للسكان تم استخدامه؟ على أي أساس تم إنشاء الطبقات؟ هل تم عرض حصص الدول واللغة الخام؟ هل طريقة تحويل الكسور إلى أعداد كاملة واضحة؟ هل تم تحديد قاعدة الانقطاع مسبقًا؟ هل احتمال الاختيار للدول الصغيرة صفر؟ هل تجمعات الدول الصغيرة منطقية ومتوافقة؟ هل تم دمج ملاحظات مراقب الدولة في الوزن العام للسكان؟ هل تم إنتاج درجة الدولة من ملاحظة مراقب الدولة؟ إذا تم أخذ عينات زائدة لمجموعات اللغة والوصول، هل تم إعادة وزنها؟ هل تم احتساب نفس الشخص كأكثر من وحدة سكانية؟ هل يتم سرد أعداد الدعوات، والمكملين، والملاحظات الصالحة بشكل منفصل؟

هل تم إنشاء قوائم النسخ الاحتياطي مسبقا؟ هل تم استبدال مشارك بسبب استجابة خاطئة أو سلبية من الذكاء الاصطناعي؟ هل تم اتخاذ قرار الصلاحية قبل تقييم دقة المواد؟ هل تم إغلاق فجوة الثابات بهدوء من دولة أخرى؟ هل تغيرت الحصة أو المجموعة بعد النتائج؟ هل يمكن إعادة إنتاج أوزان التصميم؟ هل الافتراضات حول إعداد عدم الاستجابة واضحة؟ هل تم إجراء تقليل الوزن وظهر تأثيره؟ هل يتم توفير أحجام عينات خام وفعالة معا؟ هل اختبر نفس المستخدم أكثر من منتج ذكاء اصطناعي؟ هل تم الحفاظ على هذا الاعتماد في التحليل؟ هل وصل كل منتج ذكاء اصطناعي إلى ملاحظات 1,000 على حدة؟ هل تم فصل عينات نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة؟ هل تغير إطار السكان بين الفترات؟

هل تم فصل الاتجاهات المؤطرة الثابتة والحالية؟ هل تم نشر عينات مجموعات فرعية غير كافية كما لو كانت نقاطاً نهائية؟ هل حالة جودة العينة صحيحة؟ هل حلت اللوحات التكميلية محل اللوحة الرئيسية؟ هل هناك صاحب مسؤول واضح لتسجيل العينة؟

57. الاعتراضات والردود

اعتراض 1— "لماذا لا نقدم لكل دولة الأفراد مباشرة وفقًا لنسبة السكان؟"

هذا افتراض قوي في البلدان ذات السكان الكبير والمتوسط. في الدول الصغيرة جدًا، قد تقع الحصة الخام دون شخص واحد. في هذه الحالة:

- مجموعة الدولة الصغيرة,
- الاختيار الاحتمالي للبلد،
- لوحة المراقب المنفصلة

قد تكون أكثر دقة.

الاعتراض 2— "ألن يكون من الأعدل إعطاء شخص واحد على الأقل لكل بلد؟"

قد يكون مفيدًا من حيث التغطية الجغرافية. إنه غير متناسب من حيث تقدير السكان. الحل الصحيح هو تسمية هذا الشخص بشكل منفصل كملاحظة مراقب للبلد.

الاعتراض 3— "ألا يؤدي تلقي بلد صفر أشخاص في اللوحة الرئيسية إلى استبعاده تمامًا؟"

يمكن أن تكون مشكلة إذا تلقت دولة حصة صفرية كطبقة حتمية منفصلة. يمكن إعطاء احتمال اختيار إيجابي في مجموعة الدول الصغيرة. بالإضافة إلى ذلك، يمكن لعينة المراقب الحفاظ على الرؤية الرصدية للدولة.

اعتراض 4— "هل يؤدي جمع الدول الصغيرة إلى مسح الفروقات الثقافية؟"

يمكن حذفه إذا استُخدم لتوليد درجة الدولة. بالنسبة للتقدير العالمي الرئيسي، تُدير المجموعة احتمال الاختيار. يتم الحفاظ على معرف الدولة في سجل المراقبة. يمكن عمل أخذ عينات منفصل للمراقب والدولة الخاصة.

اعتراض 5— "لماذا لا ندرج مستخدمي المراقب ضمن 1,000؟"

يمكن تضمينهم. ولكن بعد ذلك لن تكون عينة 1,000 من الأشخاص متناسبة مع السكان الأصليين. سيكون من الضروري تطبيق التوزين، وقد ينخفض حجم العينة الفعّال للتقدير الرئيسي. في الهيكل الأساسي، يتم تخصيص ملاحظات 1,000 إلى لوحة السكان؛ وتُضاف ملاحظات مراقب الدولة إلى ذلك.

اعتراض 6— "ألن ترتفع تكلفة الميدان الإجمالية كثيرًا؟"

قد يحدث ذلك. لا ينبغي أن ينتج المعيار دقة مصطنعة عن طريق إخفاء التكلفة. مع ميزانية أقل:

- عدد أقل من منتجات الذكاء الاصطناعي،
- عدد أقل من اللغات،

- حجم تجريبي أصغر،
- زيادة عدم اليقين

يمكن استخدامه. يجب كتابة الحالة الصحيحة بوضوح.

اعتراض 7— "لماذا نقوم بجمع 1,093 أو أكثر من الملاحظات عندما يوجد الاسم GEO-1000؟"

1,000 هو الهدف الصالح للملاحظة للوحة السكان الرئيسية. يمكن إضافة عينات العدالة والتغطية إلى الهدف الرئيسي. هذا لا يضر باسم المعيار. إنه يجعل حجم المجال الكلي الفعلي للقياس مرئيًا.

اعتراض 8— "لماذا يؤدي أخذ عينات مفرطة من لغة إلى تشويه النتيجة العالمية؟"

تظهر اللغة التي تم أخذ عينات مفرطة منها كبيرة لأنها في الردود الخام. عند إرجاعها إلى وزن السكان، لا يتم تشويه النتيجة العالمية. تصبح نتيجة اللغة المنفصلة أكثر دقة.

الاعتراض 9— 'الأوزان معقدة جدًا؛ أليس المتوسط الخام أسهل للفهم؟'

يمكن أن يكون المتوسط الخام ذا معنى إذا كان متناسبًا مع عينة السكان. يمكن أن تتسبب العينة المفرطة، أو عدم الاستجابة، أو اختلافات التغطية في أن يجيب المتوسط الخام على السؤال الخاطئ. لا يمكن إخفاء التعقيد إذا نشأ عن تصميم العينة.

الاعتراض 10— "إذا كان هناك 1,000 من الملاحظات الصالحة، لماذا يجب أن تكون العينة الفعالة أصغر؟"

لأن بعض الملاحظات قد تحمل وزنًا عاليًا جدًا، بينما قد تحمل ملاحظات أخرى وزنًا منخفضًا جدًا. قد تشبه السجلات من نفس العينة أو المستخدم بعضها البعض. العدد الخام هو عدد السجلات الفعلي. تقترب العينة الفعالة من حجم المعلومات الإحصائية.

الاعتراض 11— "ألن يكون من الأكثر عدلاً أن يختبر نفس الأشخاص جميع منتجات الذكاء الاصطناعي؟"

يقلل من اختلاف المستخدمين. ومع ذلك:

- يمكن أن يحد من نطاق الوصول إلى المنتج،
- يمكن أن يزيد من عبء المهام،

يمكن أن يخلق تلوثاً بين الإجابات. يمكن إدارتها باستخدام التصميم المتطابق أو تصميم الكتل الغير مكتملة المتوازن.

الاعتراض 12— "إذا لم يصل مستوى ما إلى الهدف، هل يمكن ألا يتم نشر النتيجة العالمية أبداً؟"

ليس في كل الحالات. يتم تقييم النقص من حيث:

- وزن السكان،
- المخاطر،
- سبب عدم الكفاية.
- التغطية المتبقية

تم النظر فيها. النتيجة:

- مع تغطية محدودة،
- مع عدم يقين واسع،
- مع التحذير: تغطية غير كافية

يمكن نشره. لا يمكن إخفاء النقص.

الاستئناف 13 — "إذا تغيرت بيانات السكان، هل علينا تجديد جميع الحصص في كل موجة؟"

قد يكون التجديد مطلوباً للنتيجة اللحظية الحالية. يمكن أيضاً الحفاظ على الأوزان الأولية الثابتة للمقارنة الترندية. يجب نشر النتيجتين بشكل منفصل.

الاستئناف 14 — "إذا لم نتمكن من إنشاء لوحة عالمية احتمالية، هل يمكن عدم تطبيق NOMOS؟"

إنه قابل للتطبيق. ومع ذلك، يجب تسجيل حالة جودة العينة بشكل صحيح. يمكن أن يكون مجلس من المتطوعين المعيَّرين وفقًا للسكان ذا قيمة. لا يمكن تقديمه كعينة احتمالية كاملة.

النص العام للفصل 60

لا تنشأ قوة GEO-1000 فقط من العدد .1,000 من بين 1,000 شخصًا:

- من أي سكان جاءوا،
- بأي احتمالية تم اختيارهم،
- أي دولة ولغة كانوا يمثلون،
- كم مرة تم عد نفس الشخص،
- أي الملاحظات اعتبرت صالحة،
- كيف تم التعامل مع الدول الصغيرة.
- كيف تم وزن المجموعات المفرطة العينة،
- كم عدد الأشخاص الذين لم يُكملوا المهمة،
- من تم استبداله بناءً على النتائج

تنشأ سلطة العينة من هذه الخيارات التصميمية. قد تنتج العينة المعيبة استجابات 100,000 ومع ذلك تتحدث عن السكان الخاطئين. قد تحتوي العينة الصالحة على عدد أقل من الاستجابات لكنها تدعم استنتاجًا أكثر موثوقية ضمن حدود محددة بوضوح. يجب أن يحمل سكان دولة كبيرة وزنًا كبيرًا في التجربة البشرية العالمية، لكن يجب ألا يتم اختلاق المستخدمين حيث المنتج غير متاح. يجب أن يظل وزن الدولة الصغيرة عالميًا صغيرًا، لكن لا يجب أن تكون الدولة غير مرئية. قد تمثل لغة جزءًا صغيرًا من سكان العالم بينما لا تزال تحمل خطأ خطيرًا قابلًا للقياس. قد يستخدم شخص واحد ثلاث لغات وخمسة منتجات؛ هذا لا يخلق خمسة عشر شخصًا. قد يكون رد الذكاء الاصطناعي خاطئًا دون أن يجعل المشارك غير صالح. الرد الخاطئ هو ما يحاول التدقيق قياسه.

استبداله بمستخدم جديد لا ينظف القياس. إنه يمحو الواقع. لا يجعل التوزين لوحة خاطئة تتحول معجزيًا إلى سكان العالم. ومع ذلك، يمكن أن تجعل العينات الزائدة المصممة بشكل صحيح الفئات الفرعية مرئية دون تشويه التقدير العالمي. لذلك، فإن قانون القياس السادس لـ NOMOS هو:

> يجب إنشاء العينة ليس لإنتاج النتيجة، بل لملاحظة النتيجة بعدل.

أما القانون السابع فهو كما يلي:

> لا يجب أن تزيد رؤية دولة صغيرة من وزن سكانها.

قانونها الثامن هو كما يلي:

> المراقبة الزائدة لا تعني زيادة وزن السكان.

القانون التاسع هو كما يلي:

> الإجابة الخاطئة من الذكاء الاصطناعي ليست عيبا في البيانات؛ إذا تم الالتزام به بموجب البروتوكول الصحيح، فهو مسألة رقابة.

القانون العاشر هو:

الثقة في ألف ملاحظة لا تأتي من عدد الآلاف؛ بل يأتي من قابلية تكرار الاختيار والوزن.

ترتيب الفصل 6 من NOMOS

لا تقل لي فقط: 'طلبنا *1,000* الناس.'

أظهر من بين المليارين الذين تم اختيار هذه *1,000* وبأي احتمالية.

لا تتجاهل دولة لمجرد أنها تمثل نصف شخص. / ولكن أيضًا، لا تدّع أنك قمت بقياس الدولة بأكملها لمجرد أنك أخذت عينة من شخص واحد.

اختر البلدان الصغيرة بشكل احتمالي في المجموعة المناسبة. / إذا كانت هناك حاجة للرؤية، قم بإعداد لوحة مراقب منفصلة.

يمكنك زيادة عينات اللغات قليلة الموارد. / لكن لا تجعلها أكبر في النتيجة العالمية من عدد سكانها.

لا تكرّر شخصًا بقدر الحسابات، واللغات، والجلسات، ومنتجات الذكاء الاصطناعي.

استخدم النسخة الاحتياطية عندما لا يرد المشارك. / لا تغيّر المستخدم الخاص بك عندما يعطي الذكاء الاصطناعي إجابة خاطئة.

قرر بشأن الصلاحية دون رؤية صحة إجابتك.

إذا كانت الحصة مفقودة، فاشرح ذلك. / لا تجلب شخصًا من بلد آخر لسد الفجوة.

لا تغير الحصة أو الوزن أو مجموعة الدول أو الهدف العيني بعد رؤية النتيجة.

اعرض الرقم الخام. / اعرض العينة الفعلية. / اعرض من حصل على وزن ثقيل ومن حصل على وزن خفيف.

قم بتثبيت السكان أولاً. / ثم قم بإعداد الطبقات. / ثم احسب الحصص. / ثم اختر النسخ الاحتياطية. / ثم اكتب الأوزان. / وفقط بعد ذلك، أرسل صياغة المُدخل الأولى.

جملة ختام الفصل

GEO-1000 ليس طريقة لتجميع ألف إجابة؛ إنه بروتوكول لجعل تمثيل كل شخص في السكان المستهدف قابل للقياس تحت احتمال الاختيار والوزن الصحيح.

النواة المعيارية

ينبغي أن تستهدف لوحة السكان **GEO-1000**، لكل منتج ذكاء اصطناعي ولكل موجة قياس رئيسية، ما لا يقل عن **1,000** ملاحظة صالحة للمنتج، ما لم يُبرَّر صراحة هدف مختلف. ويجب تصميم توزيع العينة وضبط إصداره وتثبيته قبل رصد استجابات الذكاء الاصطناعي. ويجب أن يستند التوزيع الأساسي إلى المجتمع المستهدف المؤهل لاستخدام المنتج، لا إلى العدد الخام للسكان المقيمين أو أعداد الدعوات أو الحسابات أو الجلسات أو سهولة التوافر. ويجب تحويل التوزيعات الكسرية إلى حصص صحيحة بطريقة معلنة مسبقًا وقابلة لإعادة الإنتاج وعمياء عن النتائج. ولا يجوز منح أي فئة ذات وزن إيجابي في المجتمع احتمال اختيار يساوي صفرًا بصورة حتمية عندما يدّعي التدقيق إمكان الاستدلال على المجتمع المستهدف كاملًا. ويجوز تمثيل الفئات السكانية الصغيرة من خلال طبقات مدمجة مبررة واختيار احتمالي متناسب. ويجوز لملاحظات مراقب الدولة والعدالة اللغوية وإمكانية الوصول أن تكمّل اللوحة الرئيسية، شريطة أن تظل موسومة بصورة منفصلة وأن يعاد ترجيحها على نحو ملائم. ولا يجوز منح المجموعات المفرط في أخذ عينات منها حصتها الخام في العينة بوصفها وزنًا سكانيًا. ويجب أن يستند استبدال المشاركين حصرًا إلى شروط تجنيد معلنة مسبقًا أو شروط تتعلق بصلاحية البروتوكول. ولا يجوز أن تؤدي استجابة ذكاء اصطناعي خاطئة أو سلبية أو رافضة أو حرجة إلى استبدال المشارك. ويجب إبقاء الأشخاص والحسابات والجلسات واللغات وملاحظات منتجات الذكاء الاصطناعي وحدات منفصلة. كما يجب توثيق أوزان التصميم وتعديلات عدم الاستجابة والمعايرة والتقليم والتجميع العنقودي والملاحظات المقترنة وحجم العينة الفعّال. وأي تغيير بعد ظهور النتائج في الحصص أو الطبقات أو المجموعات أو الأوزان أو الاستبعادات أو أهداف العينة يجب أن ينشئ إصدارًا جديدًا من التحليل مع حفظ السجل الأصلي.

الفصل 07

عينات الرصد حسب البلد والإنصاف اللغوي

الأسس المنهجية / K04; K06; K07; K20

حد الفصل

أقرت المادة 5 الحكم التالي:

> لكل منتج ذكاء اصطناعي، يجب تعريف جمهور المستخدمين المؤهلين للمنتج بشكل منفصل.

حوّل القسم 6 هذا الجمهور إلى أخذ عينات:

> يجب أن توفر لوحة السكان الرئيسية احتمالات اختيار مسبقة وأوزان للأشخاص في السكان المستهدفين قبل رؤية النتائج.

الآن نحتاج إلى حل مشكلتين مختلفتين تتعلقان بالعدالة. المشكلة الأولى:

> قد تجعل أوزان السكان البلدان الصغيرة جداً غير مرئية في العينة الرئيسية.

المشكلة الثانية:

> يمكن أن يخفي المتوسط العالمي للسكان تحريفات تمثيلية شديدة في اللغات منخفضة الموارد أو منخفضة السكان.

العينة السكانية الخالصة قوية في قياس التجربة البشرية المتوسطة في سكان العالم. ومع ذلك، نفس العينة:

- لا يمكنها مراقبة كل دولة بشكل فردي،
- لا يمكنها توفير عينات كافية لكل لغة،
- لا يمكنها إنتاج درجة للدولة للبلدان الصغيرة.

قد لا يلتقط الأخطاء النادرة لكنها الجوهرية في اللغات قليلة الموارد. لذلك، لا يمكن أن يتكون GEO-1000 من لوحة واحدة فقط. جنبًا إلى جنب مع لوحة السكان الرئيسية، هناك حاجه إلى هيكلين تكميليين:

لوحة مراقبي البلدان

ملصق اللوحة الرسمي:

لوحة مراقبي البلدان

و:

لوحة عدالة اللغة

ملصق اللوحة الرسمي:

لوحة عدالة اللغة

تجيب لوحة مراقبي البلدان على السؤال التالي:

> في البلدان التي لا يتم تمثيلها بشكل كافٍ في لوحة السكان الرئيسية، هل يوجد على الأقل مراقبة حقيقية للمستخدم، أو تحقق الوصول، أو إشارة خطر مبكرة؟

أما لوحة العدالة اللغوية، فتجيب على السؤال التالي:

> حتى لو بدا المتوسط العالمي مرتفعًا، هل يتم تمثيل المستخدمين في لغات معينة بشكل غير صحيح، أو بشكل ناقص، أو بشكل محفوف بالمخاطر؟

هاتان اللوحتان لا تحلان محل لوحة السكان الرئيسية. هما تكملانها. إعطاء مراقب دولة ملاحظة لدولة لا يقيس درجة هذه الدولة. الإفراط في أخذ عينات من لغة لا يجعل هذه اللغة أكبر في السكان العالميين. مهمة هذا القسم هي:

- رؤية الدولة مقابل وزن السكان،
- لفصل المتوسط العالمي مع عدالة اللغة،
- لفصل تقدير السكان مع الإنذار المبكر،
- لفصل الكفاية الإحصائية مع نطاق الملاحظة

عن بعضها البعض. هذا القسم لم يقم بعد بـ:

- تحديد كيفية ترجمة صياغات المُدخل في جميع اللغات.
- اختبار الترجمة العكسية والمعادلة الدلالية،
- آلية الاختيار التفصيلية للمشاركين،
- كيفية تقييم الادعاءات الذرية بواسطة المحكمين،
- حدود الامتثال النهائية

تماماً. سيتم تنظيم هذه القضايا بشكل أكبر في الأقسام التالية. السؤال الأساسي في القسم 7 هو:

> كيف نقيس أثناء إنتاج درجة عالمية دون تضخيم الدول الصغيرة واللغات بشكل مصطنع، ولكن أيضاً دون تركها غير مرئية تمامًا؟

التحدي NOMOS

لقد قمت بإعداد عينة سكانية من أفراد 1,000 لمنتج ذكاء اصطناعي. ونتيجة لعينة متناسبة مع السكان:

- مئات المستخدمين من بعض الدول الكبيرة،
- عشرات المستخدمين من بعض الدول المتوسطة الحجم،
- لا يوجد مستخدمون من معظم الدول الصغيرة

وصلوا. النتيجة العالمية:

94

تم حسابه على هذا النحو. ثم قلت: 'يمثل هذا منتج الذكاء الاصطناعي العلامة التجارية في جميع أنحاء العالم بدقة 94 بالمئة.' ومع ذلك، لا يوجد أي ملاحظة حقيقية من المستخدمين في 70 من الدول الصغيرة. قد تشكل هذه الدول جزءًا صغيرًا من سكان العالم. ومع ذلك، تبقى الأسئلة التالية بدون إجابة: هل المنتج متاح حقًا؟ هل يعمل القصد في اللغة الصحيحة؟ هل يخلط النظام بين الدولة ودولة أخرى؟ هل يتم الخلط بين الشركة المحلية والعلامة التجارية العالمية؟ هل السعر والخدمة والسياق القانوني خاطئين؟ هل يفشل النظام في الاستجابة؟ هل هناك أخطاء حرجة منخفضة التكرار؟ درجة السكان العالمية لا تجيب على هذه الأسئلة. تخيل الآن أنك أضفت شخصًا واحدًا إلى كل دولة. لقد حصلت على 70 من الملاحظات الإضافية.

لقد أضفت كل واحدة من هذه الدول إلى النتيجة الرئيسية بوزن 1/1,070، مما جعل الدول الصغيرة أكبر بكثير من حصتها الحقيقية من السكان. زاد الوضوح الجغرافي، لكن تقدير السكان كان مشوهًا. ثم تلقى مستخدم واحد في دولة واحدة ردًا غير صحيح، وصرحت: 'نتيجة GEO لهذه الدولة صفر.' رد شخص واحد ليس نتيجة لدولة. إنها

حدث واحد تمت ملاحظته في ظروف محددة في تلك الدولة. الآن اعتبر اللغة. النتيجة الإنجليزية هي 98 في المئة، الألمانية 96 في المئة، والتركية 91 في المئة. تحتوي لغة منخفضة الموارد على ثمانية مستخدمين فقط ونتيجة خام قدرها 50 في المئة. تقول، 'ثمانية أشخاص ليسوا كافيين؛ قم بإزالة اللغة من التقرير.' يرتفع المعدل العالمي، لكن تجربة المستخدمين في تلك اللغة تختفي.

في نهج آخر، تعطي مستخدمي 200 إضافيين للغة منخفضة الموارد. ينخفض المتوسط الأولي للعينة. ثم تقول: "انخفض المعدل العالمي إلى .84" ومع ذلك، قد تكون الحصة الفعلية للسكان للغة هي 2 في المئة. يزيد الإفراط في أخذ العينات القوة الإحصائية. لكنه لا يزيد من وزن السكان. الحكم الأول في هذا القسم هو كما يلي:

> أن تُلاحظ ليس هو نفسه أن يكون لديك وزن.

ينص حكمه الثاني على:

> ملاحظة واحدة لدولة ليست هي نفس درجة الدولة.

تنص الفقرة الثالثة منه على:

> الإنصاف اللغوي ليس إعطاء وزن متساوٍ لجميع اللغات في النتيجة العالمية.

تنص المادة الرابعة على:

> المتوسط العالمي العالي لا يلغي التدهور الشديد في لغة معينة.

تنص أحكامه الخامسة على:

> لا يمكن إضافة الملاحظات الإضافية التي تم أخذها لزيادة التغطية إلى النتيجة الرئيسية دون إعادة وزنها وفقًا للسكان الحقيقيين.

1. غرض الفصل

الغرض من هذا القسم هو إنشاء عينات تكميلية تزيد من رؤية الدول واللغات دون التأثير على تمثيل السكان الإحصائي في عينة السكان الرئيسية. يقوم القسم بتطبيع التمييزات التالية:

- الرؤية الجغرافية مع وزن السكان
- ملاحظات الدولة مع نتيجة الدولة
- كشف حدث واحد مع تقدير الانتشار
- اللوحة السكانية الرئيسية مع لوحة مراقب الدولة
- لوحة مراقب الدولة مع دراسة تحقق محددة لكل بلد
- تغطية الدولة مع تغطية السكان
- عدد اللغات التي يتم تغطيتها مع تغطية سكان اللغة
- عدالة اللغة مع توزيع متساوي عالمي للغات
- الحصة العالمية للغة مع أخذ عينات مفرطة
- اللغة الأم ولغة صياغة المُدخل
- اللغة والمنطقة
- اللغة ونظام الكتابة
- اللغة المدعومة واللغة التي يمكنها توليد الاستجابات بمفردها
- خلل في ترجمة المُدخل وخلل لغوي في نظام الذكاء الاصطناعي
- نقص الموارد وعيب تمثيل الذكاء الاصطناعي

- نقص المحكم والفشل في المنتج
- الاختلاف المحلي المشروع وعدم المساواة اللغوية غير العادلة
- المتوسط العالمي وقاعدة الفئة الفرعية
- تغطية الوصول إلى الدولة ودقة تمثيل الدولة
- تنبيه المراقب وقرار التوافق النهائي
- دورة مراقبة بموجة واحدة مع تغطية جغرافية متعددة الموجات

في نهاية هذا القسم، يجب أن يكون كل تدقيق GEO-1000 قادرًا على تقديم إجابات واضحة على الأسئلة التالية:

> في أي البلدان يوجد على الأقل ملاحظة فعلية واحدة؟
>
> في أي البلدان يوجد مجرد رصد للإنذار المبكر، وفي أي منها توجد توقعات إحصائية؟
>
> أي اللغات تم قياسها، وأيها وصلت إلى عينة كافية، وأيها لم يتم اختباره؟
>
> إذا تم زيادة أخذ عينات اللغات قليلة الموارد، فكيف تم إعادة وزن هذا في النتيجة العالمية؟
>
> هل مصدر النتيجة السيئة في لغة ما ناتج حقًا عن الذكاء الاصطناعي، أم أنه بسبب المشكلة في المُدخل، أو الموارد، أو الموقع، أو التحكيم؟

2. الأحكام المعيارية المركزية

> يجب على *GEO-1000* إدارة تقدير السكان الرئيسي من خلال لوحة السكان؛ ومراقبة الدول الصغيرة من خلال لوحة مراقبي الدول؛ وعدم المساواة اللغوية وتحديد اللغات منخفضة الموارد من خلال لوحة العدالة اللغوية بشكل منفصل.

هذه الهياكل الثلاثة:

- تختلف في الأغراض،
- تختلف في أهداف العينات،
- تختلف في الأوزان،
- تختلف في الادعاءات العامة

لها. لا يمكن تحويل ملاحظة من لوحة واحدة إلى معنى لوحة أخرى بدون تفسير. الهيكل الافتراضي:

العينة	الغرض الرئيسي	النتائج التي يمكن أن تنتجها	نتائج لا يمكنه إنتاجها بمفرده
لوحة السكان	تقدير عالمي موزون حسب السكان	الدرجة السكانية العالمية والمتدرجة	درجة موثوقية الدولة لكل دولة صغيرة
لوحة مراقبي البلدان	المراقبة الجغرافية والإنذار المبكر	تغطية مراقبة الدولة، حدث فردي، إشارة وصول	نسبة سكان الدولة
لوحة عدالة اللغة	التشخيص القائم على اللغة وقياس عدم المساواة	درجات اللغة، الحد الأدنى للغة، فجوة العدالة	توزيع متساوٍ للغات تلقائيًا في السكان العالميين

3. ما هو لوحة مراقب الدولة؟

لوحة مراقبي الدول، في الدول التي تمثّل تمثيلًا ناقصًا أو غير كافٍ في لوحة السكان الرئيسية:

- الوصول إلى المنتج،
- قابلية تطبيق التعليمات,
- تحليل الكيان الأساسي,
- سلوك اللغة والمنطقة,
- إشارة الخطأ الحرجة المبكرة

هي لوحة تدقيق تكميلية تم إنشاؤها للمراقبة. المهمة الأساسية للوحة مراقبي الدول ليست إنتاج درجة.

```
إنها لجعل الدولة غير المرئية مرئية لأول مرة.
```

يمكن لمراقب الدولة إظهار ملاحظة: "في البلد المحدد، وتحت منتج الذكاء الاصطناعي المحدد، واجهة المستخدم، اللغة، وظروف الوقت، تم ملاحظة هذه النتيجة مرة واحدة على الأقل." ولا يمكنه إظهار: "X٪ من المستخدمين في هذا البلد يرون نفس النتيجة."

4. وحدة مراقب الدولة

الوحدة الأساسية للملاحظة في لوحة مراقب الدولة ليست البلد بمفرده. الوحدة المفضلة:

```
S_{c,a,l,s,w}
```

حيث:

- c: الدولة أو الولاية القضائية
- a: منتج الذكاء الاصطناعي و واجهة المستخدم
- l: اللغة، نظام الكتابة، والمنطقة المحلية
- s: حالة الجلسة أو الخطة
- w: موجة القياس

هذا التمييز ضروري. في نفس البلد:

- أكثر من لغة واحدة,
- خطط مختلفة،
- واجهات الويب والجوال،
- قد توجد شروط وصول قانونية أو تقنية مختلفة

إضافة مستخدم إلى بلد معين لا تعني مراقبة جميع واجهات اللغة والمستخدمين في ذلك البلد.

5. واجبات عينة مراقب البلد

5.1. التحقق من الوصول

على الرغم من أن المنتج يبدو متاحًا في البلاد في التسجيل الرسمي، فإن المستخدم الفعلي:

- قد لا يتمكن من تسجيل الدخول،
- قد لا يتمكن من الوصول إلى الواجهة المحدد.

قد يواجه عقبة في الخطة أو التحقق. يقوم مراقب الدولة باختبارات الملاحظة على سجل الوصول الرسمي على مستوى الميدان.

5.2. فحص الهوية الأساسية

الكيان المشرف على منتج الذكاء الاصطناعي:

- الشركة الصحيحة،
- العلامة التجارية الصحيحة،
- علاقة البلد الصحيحة

هل يمكنه التحليل من الداخل؟

5.3. كشف الانحراف المحلي

في مواقع مختلفة لنفس اللغة في بلد مختلف:

- السعر،
- شركة
- الخدمة،
- القانون،
- نصيحة

هل يتغير؟ إذا كان يتغير، هل يمكن أن يكون هذا الاختلاف ماليًا وقابلًا للتفسير؟

5.4. إشارة خلل ثقيلة منخفضة التردد

في استجابة مراقب واحد:

- رخصة خاطئة،
- كيان قانوني خاطئ،
- معلومات خطرة تتعلق بالصحة أو السلامة،
- اندماج هوياتي شديد

يمكن ملاحظته. الحدث الواحد لا يدل على الانتشار في البلد. ومع ذلك، يمكن أن يحفز أعمال التحقق.

5.5. فرضية بحث جديدة

قد يطرح فريق المراقبين السؤال: 'هل هذا الحدث فردي، أم أنه نمط متكرر بين المستخدمين في البلد؟' يتم اختبار هذا السؤال بعينة إضافية مستهدفة.

5.6. سجل نطاق جغرافي

يظهر في عدد الدول المناسبة التي لدى المنتج فيها ملاحظات فعلية من المستخدمين. هذا ليس عدد درجات الدول. إنها التغطية بالملاحظة.

6. أشياء لا يمكن للوحة مراقبي الدولة القيام بها

لوحة مراقبي الدولة وحدها:

- درجة الدولة الموثوقة،
- معدل سكان الدولة،
- تصنيف الدولة،
- قرار التوافق للدولة،
- تفوق النموذج القائم على الدولة

لا يمكن إنتاجها. إذا كان هناك مراقب واحد في الدولة وكانت الإجابة صحيحة: "درجة الدولة هي 100" لا يمكن قولها. إذا كانت خاطئة: "درجة الدولة صفر" لا يمكن قولها. الحالة الصحيحة:

تم_الملاحظة_PASS

المشكلة الملحوظة

مرشح_حرج_تم_ملاحظته

فشل_الوصول

غير_كافٍ_لتقدير_الدولة

هي سجلات ملاحظة.

7. سجل الدولة المؤهلة

يجب إنشاء سجل دولة مؤهلة لكل منتج ذكاء اصطناعي. يمكن لكل دولة أن تحمل واحدة من الحالات التالية:

CE-0— غير معروف

لم يتم تحديد حالة الوصول للمنتج والتدقيق.

CE-1— خارج نطاق الوصول الطبيعي

يظل المنتج خارج المنطقة الرسمية أو الشرعية للاستخدام. لا يتم أخذه في الاعتبار لتقييم دقة تمثيل الدولة. يتم عرضه كفجوة في نطاق الوصول.

CE-2— وصول مقيد أو خاص

فقط:

- مؤسسة,
- الدعوات،
- الخطة،
- الجهاز،
- نوع المستخدم

الوصول متاح. قد تكون هناك حاجة إلى لوحة منفصلة.

CE-3— مؤهل لمراقبة مراقب الدولة

المنتج مناسب لملاحظة مراقب الدولة من حيث واجهة المستخدم ذات الصلة وعلى الأقل لغة تدقيق واحدة.

CE-4— مؤهل لتقدير السكان

لديها حصة من السكان المؤهلين للمنتج في لوحة السكان الرئيسية للبلد.

CE-5— مؤهل لتقدير البلد

يمكن إنتاج تقدير للبلد عند توفر عينة كافية، وعينة فعالة، وتغطية لغوية، وشروط جودة. يمكن أن يكون البلد CE-3 ولكنه ليس CE-5. هذا يعني أن البلد تم ملاحظته ولكن لم يمكن إنتاج درجة موثوقة للبلد.

8. السُلّم بين ملاحظة مراقب البلد وتقدير البلد

—CS-0 لم يتم ملاحظته

لا توجد ملاحظة صالحة لمراقب البلد للبلد المؤهل.

—CS-1 ملاحظة واحدة من مراقب البلد

هناك ملاحظة واحدة صالحة للدولة–المنتج–اللغة. يولّد فقط سجل حدث.

—CS-2 تكرار ملاحظة مراقب الدولة

هناك عدة ملاحظات من موجات مختلفة أو مستخدمين مختلفين. قد تثير شكوكًا حول نمط معين. غير كافية بالنسبة لمعدل الدولة.

—CS-3 عينة استكشافية للدولة

تم الوصول إلى نطاق العينة الاستكشافية في القسم 6. يمكن تقديم نتيجة استكشافية مع درجة كبيرة من عدم اليقين.

—CS-4 تقدير مؤقت للدولة

يمكن إنتاج تقدير مؤقت للدولة مع أخذ عينة كافية ومتوازنة.

—CS-5 مرشح لتقدير الدولة

هناك عينة نشطة كافية، وتغطية للغة/المكان، وتكرار للموجات، وضوابط جودة موجودة. يمكن أن يكون مرشحًا لدرجة الدولة. يجب معايرة العتبات الدقيقة من خلال تجربة أولية ومراجعة خارجية.

9. دورة المراقب

قد لا يكون مراقبة كل موجة في جميع الدول المؤهلة ممكنًا من حيث التكلفة والعمليات. لذلك، يمكن استخدام دورة المراقب. موجات القياس:

```
w=1,2,…,K
```

والدول المؤهلة: فلنعتبر CS □ c. تخصيص مراقب للدولة c في الموجة w:

```
z_{c,w}= 1(إذا كانت الدولة ستتم مراقبتها في الموجة)؛ وإلا فإن z_{c,w}= 0
```

يمكن تمثيله كما يلي. شرط الدورة الأساسية:

```
\sum_{w=1}^K z_{c,w}\geq 1
```

يجب أن يتم الوفاء به. أي يجب مراقبة كل دولة مؤهلة مرة واحدة على الأقل ضمن دورة الرصد المحددة. هذا النص:

- لا يعني أن كل دولة سيتم مراقبتها في كل موجة،
- أو أن مراقبة واحدة ستشكل نتيجة الدولة

ولا يتبع أي من الشرطين من متطلبات الدورة.

10. فئات تكرار المراقب

يمكن تقسيم الدول إلى فئات تكرار محددة مسبقًا.

SF-0— خارج النطاق

لا يوجد وصول إلى المنتج. لا يتم إجراء أي مراقبة بواسطة ممثل المستخدم. يتم مراقبة حالة الوصول.

SF-1— مرة واحدة لكل دورة مراقبة

الدول منخفضة المخاطر، الدول الصغيرة، أو الدول التي لديها بالفعل تغطية كافية غير مباشرة في لوحة السكان.

SF-2— مراقب دوري

يتم ملاحظته مرة أخرى على فترات منتظمة.

SF-3— في كل موجة رئيسية

يمكن ملاحظته في كل موجة رئيسية للأسباب التالية:

- عدد كبير من المستخدمين،
- سوق مستهدف مهم،
- خطأ جسيم سابق،
- الوصول السريع أو التغيير التنظيمي،

منطقة خدمة عالية المخاطر.

SF-4— مراقبة الحوادث

زيادة المراقبة بسبب حدث حرج تم التحقق منه أو اضطراب متواصل في التمثيل. لا يمكن تغيير فئة التكرار بأثر رجعي بعد مراجعة النتائج. يمكن لحدث جديد إنشاء نسخة جديدة للموجة التالية.

11. أولوية المراقب

يجب ألا يتم تحديد وتيرة نشاط مراقب الدولة بناءً فقط على عدد السكان. وعند الاقتضاء، يمكن أخذ المجالات التالية في الاعتبار:

- السكان المؤهلون للمنتج
- الخدمة الفعلية أو التغطية السوقية للكيان
- السجلات السابقة للتمثيل الخاطئ
- تأثير المستخدم عالي المخاطر
- التغييرات في الوصول إلى المنتج
- تفرد اللغة والموقع
- استخدام اللغة منخفضة الموارد
- الاختلافات القانونية أو التشغيلية
- الوقت المنقضي منذ آخر ملاحظة

يجب استخدام هذه الحقول وفقًا لقاعدة الحوكمة المحددة مسبقًا. نهج "نتيجة هذا البلد سيئة، لنقيس أقل" محظور.

12. مؤشرات تغطية مراقب البلد

تغطية البلد ليست رقمًا واحدًا فقط. يجب استخدام ما لا يقل عن ثلاث مقاييس منفصلة.

12.1. تغطية مراقبة الدول

نسبة الدول المؤهلة التي لديها مراقب واحد صالح على الأقل أو ملاحظة سكانية.

```
COC_count= |{c ∈ C_S: n_c^obs > 0}| / |C_S|
```

يُظهر هذا المؤشر التغطية من حيث عدد الدول.

12.2. تغطية مراقبة الدول موزونة بالسكان

نسبة السكان المؤهلين في الدول التي لديها ملاحظة واحدة صالحة على الأقل إلى إجمالي السكان المستهدفين.

```
COC_pop= [Σ_{c:n_c^obs>0} N_c] / [Σ_{c∈C_S} N_c]
```

يوضح هذا المؤشر ما إذا كانت التجمعات السكانية الكبيرة تتم مراقبتها.

12.3. تغطية التنبؤ بالبلدان

هو نسبة البلدان التي تصل إلى عينة كافية من البلدان.

```
CEC= |{c: n_{c,eff}≥ m_c}| / |C_S|
```

هنا، mc هو الحد المسبق للكفاية لتنبؤ البلد. مراجعة:

- 100 في المئة من تغطية مراقبة البلدان،
- 99 في المئة من تغطية مراقبة السكان،
- قد تحمل فقط 10 في المئة من تغطية التنبؤ بالبلدان

هذه النتائج لا يمكن تحويلها إلى بعضها البعض.

13. إنذار المراقب

عندما يتم العثور على مسألة مادية في ملاحظة المراقب، إنذار المراقب يُنشأ. أنواع الإنذارات:

SA-1— شذوذ الوصول

سجل الوصول الرسمي يتعارض مع وصول المستخدم الفعلي.

SA-2— شذوذ الكيان

الكيان يندمج مع شركة أخرى أو منتج أو عملية في دولة أخرى.

SA-3— شذوذ اللغة أو الموقع

تحدث لغة غير صحيحة، أو سياق دولة خاطئ، أو انحراف واضح في الموقع.

SA-4— شذوذ الحقائق المادية

السعر المادي، الخدمة، القدرة، الشراكة، أو الوقت غير صحيح.

SA-5— مرشح حادثة حرجة

توجد مشاكل في الصحة أو القانون أو التمويل أو الأمن أو الترخيص، أو خطأ شديد في الهوية.

SA-6— خلل في المرجع أو المُدخل

مصدر المشكلة ليس استجابة الذكاء الاصطناعي:

- مُدخل,
- الترجمة،

- المصدر الرسمي،
- التحكيم

قد يكون هو السبب. الإنذار ليس فشلًا تلقائيًا للدولة. إنه بداية التحقق والتصعيد.

14. بروتوكول تصعيد المراقب

الخطوة 1— التحقق من الصلاحية

هل الملاحظة تتوافق حقًا مع البروتوكول؟ البلد الصحيح، المنتج الصحيح، المُدخل الصحيح، اللغة الصحيحة، أول ناتج مناسب، تم التحقق من الأدلة الكاملة.

الخطوة 2— إعادة ترميز الحادث

المشكلة:

- منتج الذكاء الاصطناعي،
- سجل المرجع،
- مُدخل,
- المشارك،
- الوصول

أي من المصادر قد ينتمي إليه؟

الخطوة 3— التكرار المستقل

قبل اختيار النتيجة، يتم التحقق مع مستخدمين أو جلسات جديدة محددة مسبقًا.

الخطوة 4— أخذ عينات مستهدفة

إذا استمر الحدث، قد يتم تطبيق أخذ عينات خاص بكل بلد. حجم العينة:

- شدة الحدث،
- الانتشار المتوقع،
- عدد سكان الدولة،
- تنوع اللغة

معرَّفة مسبقًا.

الخطوة 5— فصل القرار

يتم الاحتفاظ بالنتائج التالية بشكل منفصل:

- حدث موثق واحد
- نمط متكرر
- معدل الدولة
- تأثير الملاءمة الحرج

يمكن أن يؤثر حدث حرج واحد على مراجعة المطابقة حتى بدون تقدير للانتشار.

15. ما هي العدالة اللغوية؟

العدالة اللغوية:

> الأمر لا يتعلق بأن تتلقى جميع اللغات بالضبط نفس الإجابة كلمة بكلمة.

يعني:

> في الأسئلة الموجهة للمستخدم المكافئة دلاليًا والمماثلة ماديًا، وبصرف النظر عن الفروقات المحلية الحقيقية، لا ينبغي للمستخدمين أن يتلقوا بشكل منهجي تمثيلات أكثر خطأً، أو أكثر نقصًا، أو أقل إسنادًا، أو أكثر خطورة بسبب لغتهم.

إن عدالة اللغة لا تسعى إلى هذا النوع من المساواة:

```
R_Turkish= R_English= R_Japanese
```

كلمة بكلمة. إنها تسعى إلى هذا:

> تمثيل قابل للمقارنة من حيث الهوية، والواقع المادي، والحدود، والزمن، وحالة المصدر، ومدى ملاءمته للمستخدم.

16. الفروق اللغوية ليست دائمًا ظلمًا

يمكن أن تختلف نتائج اللغة بشكل مشروع للأسباب التالية: نطاق الخدمات المحلية مختلف. تتفاوت الأسعار وشروط العقد حسب البلد. المعرفة القانونية تعتمد على الاختصاص القضائي. نية المستخدم مختلفة. التعليمات البرمجية ليست متطابقة دلاليًا. المصادر الرسمية تحتوي على محتوى مختلف عبر اللغات. منتج الذكاء الاصطناعي لا يدعم رسميًا اللغة المعنية. يجب تسجيل كل اختلاف من هذه الاختلافات. لا ينبغي تصنيف الاختلافات المحلية المفسرة كظلم.
ومع ذلك، حتى لو كانت الحقيقة نفسها، في لغة منخفضة الموارد:

- تتلف هوية الشركة،
- تُزال الحدود،
- يتم إضافة الترخيص الخطأ،
- يتم إعطاء الاستجابة باستمرار بلغة أخرى

قد يكون مشكلة في تمثيل اللغة.

17. وحدة اللغة

في تقييم اللغة، قد لا يكون مجرد اسم اللغة العام كافياً. الوحدة اللغوية المفضلة:

```
L=(λ,σ,ℓ,d)
```

ويمكن اعتبارها كذلك. هنا:

- λ: اللغة
- σ: نظام الكتابة
- ℓ: السياق المحلي أو الدولة
- d: المهمة أو المجال

على سبيل المثال، نفس اللغة:

- في نظام كتابة مختلف،

- في بلد مختلف،
- في المجال القانوني أو الصحي

قد يتطلب خلية قياس منفصلة. ليس كل مستخدمي اللغة يشكلون مجموعة متجانسة واحدة.

18. حالات دعم اللغة

يمكن لكل منتج ذكاء اصطناعي استخدام أحد الحالات التالية للغة و واجهة المستخدم:

LS-0— غير معروف

تعذر تحديد حالة الدعم.

LS-1— غير مدعوم

المنتج لا يدعم اللغة المعنية رسميًا. لا يتم تقييمه ضمن درجة اللغة المدعومة الرئيسية. قد يتم إجراء اختبارات تجريبية.

LS-2— تم ملاحظته ولكنه غير مدعوم

يمكن للمنتج توليد الردود. ومع ذلك، لا يوجد دعم رسمي أو ضمان للجودة. تم الإبلاغ عنه كنتيجة تجريبية منفصلة.

LS-3— دعم محدود

يوجد دعم لمهام أو واجهات أو ميزات محددة.

LS-4— مدعوم رسميًا

يدعي المنتج عمومًا أن اللغة مدعومة.

LS-5— مؤهل للتدقيق

اللغة:

- الدعم الرسمي،
- تكافؤ التحفيز,
- الحكم المحلي,
- مهارة المراجع,
- التقاط صالح

متوافق بالكامل مع التدقيق. وجود الدعم الرسمي وحده لا يشكل LS-5.

19. لغة منخفضة الموارد

مصطلح "لغة منخفضة الموارد" لا يصف قيمة اللغة أو جودة المتحدثين بها. في بعض سياقات التكنولوجيا والتدقيق، يشير إلى واحد أو أكثر من العيوب التالية:

- محتوى رقمي محدود
- مصدر رسمي محلي محدود
- مجموعة بيانات تقييم محدودة
- الوصول إلى المحكمين محدود

- دعم المنتج منخفض
- مصدر المصطلحات محدود
- جودة الترجمة الآلية محدودة
- وصول محدود للوحة المستخدم

يمكن أن تكون اللغة منخفضة الموارد في مجال ما وقوية في مجال آخر. على سبيل المثال، بينما تمتلك موارد قوية في المحادثة اليومية:

- القانون،
- الطب،
- مجال تقني محدد

قد تكون لديها موارد محدودة. لذلك، يجب تحديد مستوى موارد اللغة:

- المنتج,
- المهمة،
- الوقت،
- المجال

معًا.

20. مهام عينة عدالة اللغة

20.1. زيادة أخذ عينات اللغات منخفضة الموارد

يوفر مستخدمين إضافيين للغات التي لا تحصل على ما يكفي من الملاحظات في لوحة السكان الرئيسية.

20.2. قياس توزيع أخطاء المواد بناءً على اللغة

ليس فقط الدقة العامة:

- هوية
- النطاق،
- الوقت،
- المصدر،
- التوصية،
- خطأ حرج

يفحص النتائج بشكل منفصل.

20.3. التحقق من لغة الاستجابة

هل يرد الذكاء الاصطناعي بما يتماشى مع لغة التعليمات؟ إذا تحول إلى لغة أخرى، ما السبب؟

20.4. فحص التمييز بين اللغة والموقع

هل تحمل نفس اللغة واقعًا محليًا مختلفًا في بلدان مختلفة؟ هل يدير الذكاء الاصطناعي هذا الاختلاف بشكل صحيح؟

20.5. فحص التماثل بين المصدر والتمثيل الرسمي

هل تحمل السجلات الرسمية للوجود نفس الواقع المادي في جميع اللغات؟ قد ينشأ نتيجة ذكاء اصطناعي منخفضة من فجوة لغوية في المصدر نفسه.

20.6. اختبار تكافؤ صياغات المُدخل

هل تنقل الترجمات حقًا نفس نية المستخدم؟

20.7. اختبار تكافؤ المحكمين

هل يمكن للمحكمين في لغات مختلفة تطبيق نفس المعيار التنظيمي؟

21. أبواب الأربعة لعدالة اللغة

يجب تقييم نتيجة اللغة من خلال أربعة أبواب قبل نسبتها إلى أداء اللغة في منتج الذكاء الاصطناعي.

البوابة 1— تكافؤ النية

هل التعليمات قابلة للمقارنة من حيث المعنى؟ إذا فشل الأمر:

فشل تكافؤ التعليمات

تم إعطاء الحالة.

البوابة 2— تكافؤ المرجع

هل السجلات الرسمية والتحقق من الواقع كافية للغة المعنية؟ إذا فشلت:

فجوة لغة المرجع

تم إعطاء الحالة.

البوابة 3— تكافؤ القياس

هل شروط الالتقاط والجلسة والواجهة والمستخدم قابلة للمقارنة؟ إذا فشل الأمر:

فشل تكافؤ القياس

تم إعطاء الحالة.

البوابة 4— تكافؤ التحكيم

هل هناك حكام باللغة الأم أو لديهم خبرة محلية كافية ونفس قواعد التقييم؟ إذا فشل ذلك:

فجوة اللغة في التحكيم

يتم تعيينها. ما لم تتحقق المتطلبات الأربعة بشكل كافٍ: "منتج الذكاء الاصطناعي أسوأ في هذه اللغة." لا يمكن التأكيد عليه بشكل قاطع. لا تزال فجوة اللغة نتيجة ملاحظة في نظام GEO البيئي. ومع ذلك، يجب تصنيف مصدرها بشكل صحيح.

22. إنشاء الكون اللغوي

لا يمكن استخدام عبارة "جميع اللغات" دون قيود. يجب أن ينشئ كل تدقيق سجلًا لغويًا ضمن النطاق. يمكن اختيار اللغات من المجموعات التالية:

22.1. مجموعة اللغات الأساسية العالمية

اللغات الأم المستخدمة من قبل عدد كبير من البشر والمناسبة للمنتج.

22.2. مجموعة لغات سوق الكيان

اللغات المستهدفة من قبل الكيان الذي تمت مراجعته:

- الخدمة،
- من حيث المحتوى،
- العميل،
- العقد

والصيانة.

22.3. مجموعة لغات دعم المنتج

اللغات المدعومة رسميًا من قبل منتج الذكاء الاصطناعي.

22.4. مجموعة لغات المراقب منخفض الموارد

اللغات منخفضة الموارد التي تكون ضعيفة في لوحة السكان أو تشكل خطراً من حيث سلوك المنتج.

22.5. مجموعة اللغات عالية المخاطر

اللغات التي لها تأثير مادي على المستخدم في سياق الصحة أو القانون أو المالية أو الأمن.

22.6. مجموعة اللغات التجريبية

اللغات التي لا يدعمها المنتج رسميًا ولكن يتم اختبارها لأغراض المراقبة أو البحث. قد تتداخل هذه المجموعات. يجب تسجيل سبب تضمين كل لغة.

23. تخصيص عدالة اللغة

بالنسبة للغة l، ليكن الهدف في لوحة السكان الرئيسية هو nlP. الهدف الإضافي في لوحة العدالة هو nlF. الملاحظة المجمعة للغة الإجمالية:

```
n_l^T= n_l^P+ n_l^F
```

يتم الحصول عليه. يتم الاحتفاظ بالوزن الكلي للغة في حساب السكان العالمي على أنه Wl. الوزن المعمَّم لكل ملاحظة:

```
w_{i,l}= W_l/ n_l^T
```

يمكن اعتباره بهذه الطريقة. وبالتالي، يتم أخذ اللغة عينات زائدة. ومع ذلك، فإنه لا يتجاوز الحصة الفعلية للسكان في النتيجة العالمية.

24. مساهمة عينة عدالة اللغة في النتيجة العالمية

يمكن استخدام ملاحظات عدالة اللغة بطريقتين.

24.1. التحليل التشخيصي المنفصل

تنتج عينة العدالة درجات اللغة وأنواع الأخطاء فقط. المجلس السكاني الرئيسي لا يغير النتيجة العالمية. هذه الطريقة أبسط.

24.2. التحليل المشترك المعاد الوزن

تُستخدم ملاحظات عدالة السكان واللغة معًا. يتم إعادة وزن كل ملاحظة وفقًا للوزن الفعلي لسكان اللغة المستهدفة. يمكن أن تزيد هذه الطريقة من حجم المعلومات. ومع ذلك:

- العد المزدوج،
- المستخدم متعدد اللغات،
- اعتماد اللوحة،
- تباين الأوزان

يجب أخذ ذلك في الاعتبار. يجب تحديد الطريقة المستخدمة قبل رؤية النتائج.

25. مقاييس تغطية اللغة

25.1. تغطية ملاحظات اللغة

نسبة اللغات في النطاق التي لديها على الأقل ملاحظة واحدة صالحة.

```
LOC= |{l: n_l^obs > 0}| / |L_scope|
```

25.2. تغطية توقع اللغة

نسبة اللغات التي تفي بمتطلبات العينة والجودة الكافية.

```
LEC= |{l: n_{l,eff}≥ m_l}| / |L_scope|
```

25.3. تغطية اللغة بناءً على الوزن السكاني

نسبة سكان اللغة المستهدفة الذين يتم ملاحظتهم بشكل كافٍ. يتطلب وجود لغة مراجعة أساسية أو تقسيم وزن الشخص لمنع العد المزدوج للمستخدمين متعددي اللغات.

25.4. تغطية دعم المنتج

تشير إلى عدد اللغات التي يغطيها المنتج ويحصل فيها على دعم رسمي. يتم الاحتفاظ بدعم اللغة ودقة التمثيل بشكل منفصل.

26. مقاييس عدم المساواة بين اللغات

مقاييس المرشح — المقاييس المقترحة

يجب معايرة مجموعة المقاييس الدقيقة من خلال تجربة أولية ومراجعة خارجية.

26.1. نتيجة اللغة

لكل خلية لغة كافية: يتم حساب θl.

26.2. الفرق بين الأعلى والأدنى

```
DG_max= max_l θ_l- min_l θ_l
```

سهل الفهم. ومع ذلك، حساس للخلايا المفردة الصغيرة والمتقلبة.

26.3. الفرق المئوي المتين

```
DG_{90-10}= Q_{0.90}(θ_l)- Q_{0.10}(θ_l)
```

إنه أقل حساسية للخلايا القصوى.

26.4. الحد الأدنى للغة

إنه أدنى ناتج للغة مع عينة كافية.

```
LF_min= min_{l∈L_est} θ_l
```

يجب تقديم هذه القيمة مع عدم اليقين في الخلايا الصغيرة.

26.5. الحد الأدنى المتين للغة

إنه النسبة المئوية الأقل أو متوسط المجموعة الفرعية للغات الكافية. مثال:

```
LF_10= Q_{0.10}(θ_l)
```

26.6. معدل الخطأ الحرج للغة

لكل لغة:

```
CR_l= استجابة صالحة مع خطأ حرج تم التحقق منه / استجابة لغة صالحة
```

يتم نشره بشكل منفصل عن المتوسط العالمي.

26.7. معدل اللغة غير المعروفة

```
U_l= استجابة غير معروفة أو غير قابلة للحل / ملاحظة لغة ذات صلة
```

يمكن أن يكون معدل اللغة غير المعروفة العالي مهمًا بقدر أهمية النتيجة المنخفضة.

27. درجة اللغة المتساوية والدرجة الموزونة حسب عدد السكان

يمكن إنتاج نتائج تشخيصية تُعطي جميع اللغات وزناً متساويًا:

```
G_equal-language= (1/L)Σ_{l=1}^L θ_l
```

يقارب هذا الناتج السؤال: "متوسط تجربة اللغة المختبرة." إنه لا يمثل متوسط التجربة البشرية في سكان العالم.
النتيجة اللغوية الموزونة حسب عدد السكان:

```
G_population-language= Σ_l W_l θ_l
```

إنه يحافظ على توزيع السكان. الدرجتان مختلفتان. لا يمكن استخدامهما بالتبادل. يمكن عرض كلاهما معًا على بطاقة العدالة اللغوية.

28. العدالة اللغوية ليست التزامًا بتحقيق نتائج متساوية

الحصول على نفس الدرجة في لغتين ليس انتهاكًا تلقائيًا. يمكن تفسير الفرق للأسباب التالية:

- فرق الخدمة المحلية الفعلي
- القانون الخاص بالدولة
- تغطية الدعم الرسمي للمنتج
- الوصول إلى الموارد
- نية المستخدم
- منطقة الاستدعاء

ومع ذلك، إذا كان مصدر الفرق يمكن تفسيره أو كانت اللغات ذات الموارد المنخفضة تمر بشكل منهجي:

- هوية خاطئة،
- تفويض غير صحيح،
- نقص المادة،
- خطأ حرج عالي

إذا استمر، فهناك مشكلة في عدالة اللغة. NOMOS يعرف عدالة اللغة كما يلي:

> في ظروف مادية مكافئة، لا ينبغي أن تكون اللغة نفسها سببا للتمثيل الخاطئ المنهجي.

29. تقاطع اللغة والبلد

درجة اللغة ليست مستقلة عن الدول. نفس اللغة:

- دول مختلفة،
- الوصول إلى منتجات مختلفة،
- الموارد المحلية المختلفة،
- القانون المختلف،
- السعر المختلف

يمكن استخدامه في سياق. لذلك، يجب فحص خلية البلد–اللغة: (c,l) بشكل منفصل عند الضرورة. مثال:

- ألمانيا × التركية
- تركيا × التركية
- المملكة المتحدة × التركية

إنها نفس اللغة ولكن قد تحمل سياقًا محليًا وتجاريًا مختلفًا. لا يمكن قياس عدالة اللغة فقط من خلال اللغة الرسمية للبلد.

30 المشاهد الوطني AND عينة عدالة اللغة التقاطع

ملاحظة في وقت واحد:

- ملاحظة مراقب دولة صغيرة،
- ملاحظة لغة منخفضة الموارد

قد يحدث. في هذه الحالة، يجب أن يحتوي السجل على الحقول التالية:

الدور الرئيسي في اللوحة

العلامات التشخيصية الثانوية

وزن السكان حالة المراقب حالة عدالة اللغة لا يمكن احتساب نفس الملاحظة بالكامل في لوحتين منفصلتين. إنها ملاحظة واحدة. لديها أكثر من علامة تشخيصية.

31. تغيير المشاركين في لوحة المراقب واللغة

يمكن تغيير المشارك فقط بسبب انتهاك البروتوكول. النتائج التالية ليست أسباباً للتغيير:

- استجابة الذكاء الاصطناعي غير الصحيحة
- الرفض

- الاستجابة بلغة أخرى
- خطأ حرج
- ارتباك الكيان
- توصية سلبية

هذه هي نتائج القياس. الغرض من لوحة المراقبين هو تحديد الأحداث غير المتوقعة بشكل محدد. تغيير المستخدم لأن حدثًا غير متوقع حدث يُفشل الغرض من اللوحة.

32. تحكيم اللغة

يجب أن يستوفي التحكيم في عينة عدالة اللغة على الأقل الشروط التالية:

- القدرة على فهم المُدخل والاستجابة بشكل طبيعي أو بمستوى عالي من الكفاءة
- معرفة المصطلحات المحلية
- القدرة على تقييم مصادر الأدلة
- باستخدام نفس دليل التقييم المعياري
- كونك أعمى عن اهتمامات النموذج وإذا أمكن اهتمامات العلامة التجارية
- الإفصاح عن تضارب المصالح
- مسار الحكم الثالث في حالة الخلاف

لغة لا يمكن العثور على محكم لها:

فجوة اللغة في التحكيم

يجب أن يكون له الحالة. لا يمكن الافتراض أن جميع الفروق الدقيقة صحيحة إذا تم ترجمتها إلى الإنجليزية بواسطة الترجمة الآلية. يمكن أن تكون الترجمة الآلية أداة مفيدة. لكنها ليست التحقق النهائي من المعنى المحلي.

33. توازن المصدر

عن الكيان في الموقع الإنجليزي:

- النطاق،
- السعر،
- رخصة،
- حدود,
- دليل

موجود؛ إذا لم تكن هذه موجودة في نسخة لغة أخرى، فإن النتيجة المنخفضة للذكاء الاصطناعي قد لا تكون مجرد مشكلة نموذجية. هناك حاجة إلى ثلاث نتائج منفصلة:

- الكيان اللغة المصدر الفجوة
- المنتج الذكاء الاصطناعي اللغة تمثيل الفجوة
- الفجوة المجمعة التي يواجهها المستخدم

المشكلة النهائية من منظور المستخدم لا تزال حقيقية. ومع ذلك، قد تختلف مسؤولية التصحيح:

- الكيان،
- مزود الذكاء الاصطناعي،

كلاهما.

34. الادعاء بالتكافؤ

صياغات المُدخل التي ليست متكافئة دلاليًا تنتج إجابات مختلفة. مثال: في لغة واحدة: "ماذا تفعل هذه الشركة؟" في لغة أخرى: "هل هذه الشركة موثوقة وتستحق التوصية؟" إذا تم السؤال، لا يمكن مقارنة النتائج. سيتم تفصيل دستور صياغات المُدخل في القسم التالي. نص هذا القسم كما يلي:

> لا يمكن نسب اختلافات أداء نموذج اللغة إلى أداء الذكاء الاصطناعي اللغوي دون التحقق من تكافؤ صياغات المُدخل.

35. انحراف لغة الاستجابة

يقوم المستخدم بإعطاء مطالبة باللغة التركية. ينتج الذكاء الاصطناعي إجابة باللغة الإنجليزية. هذه النتيجة:

- حتى لو كانت المعلومات الواقعية صحيحة،
- قد تؤثر على تجربة المستخدم.
- سهولة الوصول،
- والفهم.

يجب تسجيل انحراف لغة الاستجابة بشكل منفصل:

اللغة المطابقة

تبديل الشفرات الجزئي

تبديل اللغة غير المطلوب

إخراج اللغة غير قابل للاستخدام

مجهول

تبديل الشفرات ليس دائمًا خطأً. على سبيل المثال، قد يكون استخدام المصطلحات التقنية باللغة الإنجليزية أمرًا طبيعيًا. ومن المهم إذا كان يعطل إمكانية استخدام الاستجابة الرئيسية.

36. نظام الكتابة والنقل الصوتي

يمكن استخدام نفس اللغة في أنظمة كتابة مختلفة. بالنسبة لـ:

- الاسم،
- اسم المنتج،
- الهوية القانونية

قد يتم الخلط بينه وبين كيان آخر أثناء النقل الصوتي. عينة عدالة اللغة، بقدر ما هو ذي صلة:

- نظام الكتابة،
- النقل الصوتي،
- الاسم المحلي الرسمي،

- الاسم المستخدم من قبل المستخدم

يجب عليهم تسجيل حقولهم. يجب اعتبار اختلافات نظام الكتابة رسمية فقط. إذا أثر ذلك على تحديد الكيانات، فهو مادي.

37. مراقب الدولة التخليقي CYCLE

عرض المنهجية الاصطناعية / المنتج التالي والدول والأرقام والنتائج كلها خيالية. في سجل المنتج الاصطناعي المسمى Orion AI نطاق الوصول الطبيعي: هناك 120 دولة مؤهلة. وقد أنتج اللوح السكاني الرئيسي مراقبات صالحة في الموجة الأولى من 38 دول فقط بسبب وزن السكان.

37.1. نطاق الموجة الأولى

عدد الدول المؤهلة: 120 الدول التي تحتوي على مراقبة واحدة على الأقل: 38

```
COC_count= 38/120= 31.7%
```

دع هذه 38 الدول تمثل 97% من السكان المؤهلين.

```
COC_pop= 97%
```

النتيجة: عدد الدول منخفض في التغطية، بينما تغطية السكان مرتفعة.

37.2. دورة مراقب الأربع موجات

وزّع الدول المتبقية 82 على أربع موجات باستخدام برنامج عشوائي قائم على المخاطر ومجمد سابقًا. يتم مراقبة كل دولة مرة واحدة على الأقل خلال الدورة. في نهاية الموجة الرابعة:

```
COC_count= 120/120= 100%
```

ومع ذلك، تم الوصول إلى عينة كافية لتقدير الدول في 14 دول فقط.

CEC = 14/120= 11.7%

السجل العام الصحيح: تغطية ملاحظات الدولة: /100% تغطية ملاحظات السكان: بالقرب من /100% تغطية توقعات الدولة: 11.7% السجل العام غير الصحيح: "جميع دول 120 لديها درجة دولة موثوقة."

37.3. إنذار حرج لدولة صغيرة

في سجل مراقب الدولة، الملاحظة المستخلصة من استجابة الذكاء الاصطناعي في الدولة الصغيرة **J** تُخطئ في تحديد الكيان المدقق كمكتب محاماة مرخص. في الواقع، الكيان لا يقدم أي خدمات قانونية. الملاحظة الواحدة لا تُقدّر معدل الخطأ في الدولة **J**؛ إنها تحدد حدثًا حرجًا محتملًا. البروتوكول: تحقق من الملاحظة؛ تحقق من المُدخل والسجل المرجعي؛ كرر مع مستخدمين مستقلين جدد؛ إذا استمر الحدث، افتح عينة مستهدفة من الدولة؛ وقيم تأثير الامتثال الحرج بشكل منفصل. الحدث الملحوظ قد يكون نادرًا؛ الضرر المحتمل لا يجب أن يكون صغيرًا.

38. عرض عدالة اللغة الاصطناعية

عينة اصطناعية — ليست أداء حقيقي للذكاء الاصطناعي أو الشركة

تضمين خمس لغات في اختبار الهوية الأساسية لـ Apple.com Synthetic. تخصيص اللوحة السكانية الرئيسية:

اللغة	وزن السكان	ملاحظة اللوحة الرئيسية
L1	55%	550
L2	25%	250

اللغة	وزن السكان	ملاحظة اللوحة الرئيسية
L3	10%	100
L4	6%	60
L5	4%	40
الإجمالي	100%	1,000

نتائج اللغة الاصطناعية:

اللغة	معدل النجاح
L1	95%
L2	92%
L3	88%
L4	65%
L5	58%

النتيجة العالمية المرجحة حسب السكان:

```
0.55(95)+0.25(92)+0.10(88)+0.06(65)+0.04(58)=90.27
```

النتيجة العالمية مرتفعة. ومع ذلك، هناك تدهور شديد في لغتين.

38.1. زيادة العينات للمساواة بين اللغات

إضافة ملاحظات إضافية من 140 إلى 160 و L4 إلى L5. الأعداد الجديدة للملاحظات:

اللغة	إجمالي الملاحظات
L1	550
L2	250
L3	100
L4	200
L5	200
الإجمالي	1,300

المتوسط الخام عبر جميع ملاحظات 1,300 أقل من النتيجة المرجحة حسب السكان الموضحة أعلاه. لا يُعد هذا تقييمًا للسكان العالميين، لأن L4 و L5 تم تكثيف أخذ عيناتهم عمدًا. بمجرد استعادة أوزان السكان المستهدفة، تعود النتيجة العالمية إلى الرقم المرجح حسب السكان أعلاه. يساعد التكثيف على تحسين الحساسية لـ L4 و L5 دون تغيير التقدير المرجح حسب السكان.

38.2. بطاقة تقييم إنصاف اللغة

نتيجة تركيبية. النتيجة العالمية المرجحة حسب السكان: 90.27. درجة التساوي بين اللغات:

```
(95+92+88+65+58)/5=79.6
```

أعلى إلى أدنى فرق لغوي: 37 نقاط؛ أقل لغة: 58؛ متوسط أدنى لغتين: 61.5؛ معدل الأخطاء الحرجة في اللغات ذات الموارد المنخفضة: يُبلَغ عنه بشكل منفصل؛ تكافؤ صياغات المُدخل: يفترض أنه تم التحقق منه؛ تكافؤ المرجع: يُفترض أنه تم التحقق منه. التفسير الصحيح:

يبدو أن تمثيل السكان العالمي مرتفع؛ ومع ذلك، العدالة اللغوية ضعيفة بشكل كبير.

تفسير غير صحيح: "نظرًا لأن النتيجة العالمية هي 90، فإن GEO ناجح في جميع اللغات."

39. حالة اصطناعية تميز مصدر تدهور اللغة

لننظر في ثلاث لغات.

اللغة أ

معادل صياغة المُدخل المصدر الرسمي كامل الحكم المحلي كافٍ استجابة الذكاء الاصطناعي غير صحيحة النتيجة: مرشح لمشكلة تمثيل اللغة في منتج الذكاء الاصطناعي.

اللغة ب

ترجمة الفوري تختلف جوهريًا الاستجابة الذكائية الاصطناعية مناسبة للفوري النتيجة: خلل في دستور الفوري.

اللغة ج

في صفحة اللغة الرسمية، يتم نقل السعر القديم من المصدر الإنجليزي لـ AI إلى السعر القديم الحالي: مشكلة توازن مصدر لغة الكيان والفجوة في التمثيل المشترك التي تصل إلى المستخدم. حتى إذا ظهرت درجات اللغات الثلاث متساوية، فإن مالك التصحيح مختلف.

40. لوحة الدولة واللغة بطاقة النتيجة العامة

يجب أن تتضمن بطاقة النتيجة العامة للمرشح الحقول التالية:

حقول الدولة

نطاق الوصول الطبيعي العدد المناسب من الدول عدد الدول التي لديها ملاحظة واحدة على الأقل تغطية الملاحظة للدولة تغطية الملاحظة الموزونة حسب السكان تغطية التنبؤات للدولة وقت دورة المراقب للدولة تنبيهات المراقب المفتوحة الدول خارج الوصول الأصلي الدول غير المرصودة تاريخ القياس

حقول اللغة

عدد اللغات المغطاة عدد اللغات المرصودة عدد اللغات التي يمكن عمل تنبؤات لها اللغات المدعومة رسميًا اللغات التجريبية تغطية نتيجة اللغة العالمية موزونة حسب السكان نتيجة الاعتراف المتساوي باللغات الحد الأدنى للغة فرق اللغة معدلات الخطأ الحرجة

فجوة لغة المرجع

فشل تكافؤ التعليمات

لغات غير مختبرة نسخة لوحة اللغة تُظهر هذه البطاقة الواقع الذي لا يمكن لشارة واحدة "متعددة اللغات" تحمله.

41. الأحكام الإلزامية المعيارية

CH07-N01

يجب تعريف لوحات مراقب الدولة وعدالة اللغة بشكل منفصل عن لوحة السكان الرئيسية من حيث الغرض والعينة والأوزان.

CH07-N02

لا يمكن استخدام ملاحظة مراقب الدولة كدرجة سكان الدولة.

CH07-N03

بالنسبة للدولة، وجود ملاحظة والقدرة على إنتاج تقدير للدولة هما حالتان منفصلتان.

CH07-N04

يجب الإبلاغ عن عدد الدول المشمولة وتغطية الدول وفقًا للوزن السكاني معًا.

CH07-N05

يجب عرض تغطية تقدير الدولة بشكل منفصل عن تغطية الملاحظات.

CH07-N06

يجب أن تمتلك كل دولة مؤهلة احتمالية للملاحظة أكبر من صفر في دورة المراقبة المحددة؛ يجب أيضًا عرض الدول غير المشمولة بالنطاق.

CH07-N07

يجب أن تكون دورة المراقبة وفئات تكرار الدول مرقمة قبل عرض استجابات الذكاء الاصطناعي.

CH07-N08

لا يمكن إزالة دولة منخفضة الدرجة من دورة المراقب بعد رؤية النتائج.

CH07-N09

تنبيه مراقب الدولة ليس فشلًا تلقائيًا للدولة؛ هو يحفز التحقق والتصعيد.

CH07-N10

يمكن لحدث مراقب حاسم واحد أيضًا أن يبدأ مراجعة امتثال منفصلة دون تقدير الانتشار.

CH07-N11

لا يمكن استخدام الملاحظات من عينة مراقب الدولة في العينة السكانية الرئيسية بالوزن المتساوي الخام.

CH07-N12

إذا تم تضمين ملاحظات مراقب الدولة في النتيجة الرئيسية، يجب نمذجة احتمالات اختيارها وأوزانها بشكل صريح.

CH07-N13

لا يمكن تعريف العدالة اللغوية على أنها الالتزام بإعطاء نفس الإجابة كلمة بكلمة.

CH07-N14

يتطلب العدالة اللغوية دقة مشابهة وحدود ومصدر وملاءمة للمستخدم تحت نفس الظروف المادية.

CH07-N15

اللغة والبلد، اللغة والموقع، واللغة ونظام الكتابة لا يمكن استخدامها بالتبادل.

CH07-N16

يجب شرح الكون اللغوي المغطى وسبب تضمين كل لغة قبل القياس.

CH07-N17

يمكن استخدام ادعاء 'جميع اللغات' فقط إذا تم بالفعل قياس جميع الواجهات اللغوية والمستخدم ذات الصلة؛ وإلا، يجب قول 'اللغات ضمن النطاق'.

CH07-N18

لا يمكن إضافة الملاحظات بلغات غير مدعومة رسميًا إلى أداء المنتج الرئيسي المدعوم دون توضيح.

CH07-N19

يجب ألا يؤدي الإفراط في أخذ عينات اللغات ذات الموارد المحدودة إلى زيادة الوزن السكاني العالمي.

CH07-N20

إذا تم تضمين ملاحظات من عينة عدالة اللغة في النتيجة العالمية، فيجب إعادتها إلى أوزان سكان اللغة الهدف الفعلية.

CH07-N21

لا يمكن احتساب نفس المستخدم متعدد اللغات كأكثر من إنسان واحد كامل في وزن السكان العالمي.

CH07-N22

دون التحقق من تكافؤ المطالب، لا يمكن عزو الفرق في درجة اللغة بشكل قاطع إلى أداء اللغة لدى الذكاء الاصطناعي.

CH07-N23

لا يمكن تحميل درجة اللغة المنخفضة على منتج الذكاء الاصطناعي وحده دون تقييم التكافؤ المرجعي.

CH07-N24

لا يمكن اتخاذ قرار بشأن دقة اللغة بشكل نهائي دون وجود كفاءة كافية لدى الحكّام المحليين.

CH07-N25

يجب تسجيل انحراف لغة الاستجابة وتبديل الرموز بشكل منفصل.

CH07-N26

لا يمكن تصنيف الاختلافات الحقيقية المشروعة محليًا على أنها ظلم لغوي.

CH07-N27

لا يمكن تجاهل الفروق الملموسة الغامضة في اللغة والاكتفاء باعتبارها مجرد فروق تخص التخصيص أو الأسلوب.

CH07-N28

لا يمكن لنقاط السكان العالميين أن تحل محل نتيجة عدالة اللغة.

CH07-N29

لا يمكن تقديم درجة اللغة المتساوية مثل الدرجة العالمية الموزونة حسب السكان.

CH07-N30

يجب عرض معدلات الأخطاء الحرجة القائمة على اللغة بشكل منفصل عن المتوسط العام.

CH07-N31

يجب الحفاظ على PROMPT_EQUIVALENCE_FAILURE وREFERENCE_LANGUAGE_GAP وMEASUREMENT_PARITY_FAILURE وADJUDICATION_LANGUAGE_GAP كحالات منفصلة.

CH07-N32

في عينات مراقب الدولة وعدالة اللغة، لا يمكن أن يستند قرار تغيير المشاركين على صحة استجابة الذكاء الاصطناعي أو لفائدة العلامة التجارية.

CH07-N33

يمكن أن يكون للملاحظة نفسها عدة تسميات تشخيصية؛ ومع ذلك، لا يمكنها الحصول على وزن كامل أكثر من مرة واحدة في التنبؤ العالمي دون تفسير.

CH07-N34

يجب أن تحتوي سجلات لوحة الدولة واللغة على مالك بشري أو مؤسسي مسؤول.

42. أشكال الفشل

CH07-F01— نتيجة الدولة لشخص واحد

يتم نشر ملاحظة مراقب واحد كنتيجة GEO للدولة.

CH07-F02— تغطية الدولة الرمزية

يتم تعيين شخص واحد لكل دولة ويطلق عليه 'متعلقاً بالسكان العالميين.'

CH07-F03— الخلط بين تغطية السكان وعدد الدول

على الرغم من ملاحظة عدد قليل من الدول الكبيرة، إلا أنه يُعرض فقط تغطية السكان؛ وعلى الرغم من ملاحظة عدد كبير من الدول الصغيرة، إلا أنه يُعرض فقط عدد الدول.

CH07-F04— احتساب نطاق الملاحظة كنطاق تقديري

تُعتبر جميع الدول التي لديها على الأقل ملاحظة واحدة دولًا مُسجلة.

CH07-F05— احتساب ملاحظة مراقب دولة مرتين في النتيجة الرئيسية

تحصل نفس الملاحظة على الوزن الكامل في كل من لوحة المراقب ولوحة السكان.

CH07-F06— تحويل تنبيه مراقب الدولة إلى ادعاء بالانتشار

يتم تعميم حدث واحد على جميع المستخدمين في الدولة.

CH07-F07— تجاهل حدث المراقب الحرج

يتم تجاهل المخاطر العالية عن طريق قول "مستخدم واحد فقط".

CH07-F08— إزالة دولة منخفضة التقييم من الدورة

يتم تغيير جدول المراقب وفقًا للنتائج.

CH07-F09— مراقبة الأسواق التجارية فقط

يتم استبعاد البلدان التي لديها وصول إلى المنتجات ولكن تعتبر غير مهمة تجاريًا تمامًا؛ يتم عرض النتيجة عالميًا.

CH07-F10— اعتبار الدولة واللغة متطابقين

تُقبل اللغة الرسمية للبلد كلغة لجميع المستخدمين.

CH07-F11— جميع اللغات ادعاء

يتم قياس عدد قليل من اللغات المحدودة؛ تُنشر النتيجة على أنها "في جميع اللغات".

CH07-F12— مسح لسان الموارد المنخفضة

نظرًا لصغر العينة، يتم استبعاد اللغة تمامًا من التقرير.

CH07-F13— نمو مفرط للغة منخفضة الموارد على المستوى العالمي

تدخل اللغة التي تم أخذ عينات زائدة عنها في النتيجة العالمية بحصتها الخام من العينة.

CH07-F14— احتساب درجة اللغة المتساوية باعتبارها درجة السكان

يتم تقديم النتيجة التشخيصية التي تعطي وزنًا متساويًا لكل لغة على أنها متوسط تجربة سكان العالم.

CH07-F15— احتساب درجة السكان باعتبارها عدالة لغوية

النتيجة العالمية، المدعومة بوزن اللغات الكبيرة، تخفي التشويه في اللغات الصغيرة.

CH07-F16— احتساب خطأ ترجمة الطلب كخطأ للذكاء الاصطناعي

يتم نسب الاختلاف الناتج عن الطلبات ذات المعاني المختلفة إلى النموذج.

CH07-F17— احتساب فجوة اللغة المصدر كخطأ للذكاء الاصطناعي

على الرغم من عدم وجود معلومات رسمية محلية، يتم إعطاء كل المسؤولية لمنتج AI.

CH07-F18— الدفاع عن خطأ الذكاء الاصطناعي بنقص الموارد

حتى مع توفر نفس الموارد الكافية، يتم إغلاق الأخطاء في اللغة منخفضة الموارد على أنها محتوى مفقود فقط.

CH07-F19— اعتبار نقص لغة المحكم فشلًا في المنتج

حتى لو أساء المحكم الفهم، يتم اعتماد النتيجة نهائيًا.

CH07-F20— اعتبار الترجمة الآلية كمحكم نهائي

لا يتم التحقق من فقدان المعنى المحلي.

CH07-F21— اعتبار تغييرات الكود خطأ تلقائي

تحدث الأخطاء مع المصطلحات التقنية الطبيعية أو خلط اللغات المقبول.

CH07-F22— اعتبار الانحراف اللغوي غير القابل للاستخدام أسلوبًا

حتى إذا استجاب النظام بلغة أخرى لا يستطيع المستخدم فهمها، يُعتبر الناتج ناجحًا.

CH07-F23— تجاهل اختلاف نظام الكتابة

يعتبر خطأ التحويل الصوتي الذي يسبب الالتباس بين الكيانات خطأً رسميًا.

CH07-F24— احتساب مستخدم متعدد اللغات مرتين

يتم احتساب نفس الشخص كوحدة سكانية منفصلة في جميع اللغات.

CH07-F25— اعتبار الاختلاف المحلي المشروع غير عادل

حتى إذا كان الاختلاف في السعر أو القانون قائمًا على ظروف محلية فعلية، يُعلن على أنه تحيز لغوي.

CH07-F26— اعتبار الاختلاف غير المفسر شخصيًا

عدم التناسق المادي في ظروف معادلة لا يُشكك فيه.

CH07-F27— النتيجة المحددة من خلية اللغة غير الكافية

يتم تصنيف اللغة ببعض الملاحظات.

CH07-F28— ضبط لوحة اللغة بعد النتيجة

يتم اختيار اللغات التي حصلت على درجات منخفضة أو عالية فقط بعد ذلك.

CH07-F29— قياس لغات المصدر العالية فقط

يتم تقديم النتيجة على أنها "متعددة اللغات".

CH07-F30— الحفاظ على اللغات غير المختبرة

لا يظهر حقل غير المختبر على بطاقة النتيجة العامة.

43. إجراء التدقيق

الخطوة —1 تجميد سجل الامتثال المنتج-البلد

لكل بلد:

- نطاق الوصول الطبيعي
- الوصول للمنتج
- لغة التدقيق
- واجهة المستخدم

يتم تسجيل الحالة.

الخطوة —2 استخراج تغطية البلدان في اللوحة السكانية الرئيسية

يتم تحديد البلدان التي توجد فيها الملاحظة في اللوحة الرئيسية.

الخطوة —3 تحديد الكون الخاص ببلدان المراقب

يتم تحديد البلدان التي تعاني من نقص في اللوحة الرئيسية ولها حالة CE-3.

الخطوة —4 إنشاء دورة المراقب

البلدان:

- تُخصص من حيث فئة التكرار،
- موجة,
- اللغة/الإعدادات المحلية,
- واجهة المستخدم

والصيانة.

الخطوة —5 تحديد المشاركين كمراقبين مسبقًا

يتم إنشاء القوائم الرئيسية والاحتياطية قبل رؤية النتائج.

الخطوة —6 تصنيف نتائج المراقب

يتم تعيين الوصول، الهوية، اللغة، الحقائق المادية، وحالات الأحداث الحرجة.

الخطوة —7 تطبيق الإنذار والتصعيد

حسب نوع الإنذار:

- مرة أخرى,
- عينة مستهدفة
- مراجعة المراجع
- مراجعة الامتثال

تم البدء.

الخطوة —8 حساب مقاييس تغطية البلدان

يتم حساب نطاق عدد البلدان، ونطاق السكان، ونطاق تقدير البلد بشكل منفصل.

الخطوة —9 تجميد سجل اللغة

لكل لغة:

- لأسباب النطاق،
- حالة الدعم،
- نظام الكتابة،
- المكانية,
- كون المستخدمين

يتم تسجيلها.

الخطوة —10 إنشاء أخذ عينات إضافية لضمان عدالة اللغة

يتم وضع أهداف إضافية للغات التي تكون غير كافية في لوحة السكان الرئيسية.

الخطوة —11 التحقق من البوابات الأربع للغة

تقييم التكافؤ النظامي، تكافؤ المرجع، تكافؤ القياس، تكافؤ التحكيم.

الخطوة —12 ملاحظات لغة النتيجة

لكل لغة:

- الصدق،
- نقص المادة،
- خطأ حرج
- غير معروف،
- انحراف لغة الاستجابة

يتم تسجيلها.

الخطوة —13 فصل الأوزان العالمية والتشخيصية

يتم حساب النتيجة الموزونة حسب السكان والنتيجة مع الاعتراف المتساوي باللغة بشكل منفصل.

الخطوة —14 حساب عدم المساواة اللغوية

أرضية اللغة، فرق لغة ثابت، فرق خطأ حرج، يتم حساب معدل المجهول.

الخطوة —15 فحص تقاطعات الدولة–اللغة

يتم تحديد الاختلافات أو الصراعات المحلية للمواد.

الخطوة —16 إنشاء بطاقة نتائج عامة

يتم إظهار النطاق، كفاية التنبؤ، التنبيهات، والمناطق غير المختبرة.

44. الدليل الضروري

سجل المنتج والواجهة للمستخدم للذكاء الاصطناعي؛ سجل الدولة نطاق الوصول الطبيعي؛ عدد الدول المؤهلة؛ حالة الوصول إلى الدولة؛ توزيعات الدولة الرئيسية للوحة السكان؛ كون المراقب للدولة؛ دورة المراقب؛ فئات تكرار المراقب؛ مصفوفة تعيين الدولة–الموجة؛ تعيينات الدولة–اللغة؛ قوائم المراقبين الأساسية والاحتياطية؛ نتائج مراقبة الدولة؛ تنبيهات المراقب؛ سجلات التصعيد؛ عينات الدول المستهدفة؛ تغطية مراقبة الدولة؛ تغطية الدولة الموزونة

بالسكان؛ تغطية تقدير الدولة؛ سجل اللغة ضمن النطاق؛ حالة دعم اللغة؛ سجلات النصوص واللوكال; أعداد عينات اللغة الرئيسية؛ أعداد تجاوز العدالة اللغوية في العينات؛ أوزان اللغة حسب السكان. سجلات المستخدم متعدد اللغات؛ سجلات تكافؤ الحافز؛ سجلات تكافؤ المرجع؛ سجلات تكافؤ القياس؛ سجلات تأهيل المحكم المحلي؛ ونتائج اللغة.

اللغة معدلات الأخطاء الحرجة انحرافات لغة الاستجابة مقاييس عدم المساواة بين اللغات اللغات غير المختبرة نتائج متساوية للغة والسكان أوزان العينة فحوصات العد المزدوج النتائج العامة البطاقة إصدارات العينة الشخص أو المؤسسة المسؤولة

45. قائمة تدقيق التدقيق

هل تم فصل هدف لوحة المراقب الوطني عن لوحة السكان الرئيسية؟ ما الدول المدرجة في المراقب؟ هل حالتا الأهلية والوصول للدول محدثة؟ هل هناك احتمال لملاحظة إيجابية في دورة المراقبة لكل دولة مؤهلة؟ هل تم تجميد دورة المراقبة قبل النتائج؟ هل تم استخدام ملاحظة المراقب الوطني مثل تقييم الدولة؟ هل تم فصل تغطية مراقبة الدولة وتغطية تقدير الدولة؟ هل تم عرض تغطية الدولة الموزونة حسب السكان بشكل منفصل؟ هل تم احتساب ملاحظات المراقب الوطني مرتين في التقييم الرئيسي؟ هل تم تأكيد تنبيه المراقب الحرج؟ هل تم تحويل إنذار واحد إلى ادعاء بالانتشار؟ هل تم فتح أخذ عينات مستهدف بسبب الإنذار؟ هل تم تحديد كون اللغات في نطاق الدراسة مسبقًا؟ هل تتوافق ادعاء 'جميع اللغات' مع التغطية الفعلية؟ هل تم فصل اللغة ونظام الكتابة والمحلية؟

هل تم تسجيل حالة دعم اللغة للمنتج؟ لماذا تم اختيار اللغات قليلة الموارد؟ هل تم إعادة وزن أمثلة عدالة اللغة في النتيجة العالمية؟ هل تم احتساب نفس المستخدم متعدد اللغات مرتين؟ هل تم التحقق من معادلة صياغات المُدخل؟ هل المصادر الرسمية متساوية جوهريًا عبر اللغات؟ هل يعرف الحكام اللغة المعنية بمستوى كافٍ؟ هل تم استخدام الترجمة الآلية كحكم نهائي؟ هل تم تسجيل انحرافات لغة الاستجابة؟ هل تم فصل التباين المحلي المشروع عن ظلم اللغة؟ هل تم نشر مجموعة بيانات اللغة وتباين اللغة؟ هل معدلات الأخطاء الحرجة القائمة على اللغة مرئية؟ هل تم فصل فجوة لغة المرجع REFERENCE_LANGUAGE_GAP وغيرها من عيوب القياس؟ هل اللغات غير المختبرة مرئية؟ هل يخفي المقياس العام مشكلة عدالة اللغة؟

هل حصلت نفس الملاحظة على كامل الوزن عبر عدة عينات؟ هل تم تحديد صاحب المسؤولية عن سجلات العينة؟

46. الاعتراضات والردود

الاعتراض 1— «أليس من الأفضل إعطاء شخص واحد لكل دولة صغيرة بدلاً من عدم إعطاء أي شيء؟»

للمراقبة المبكرة والتغطية الجغرافية، نعم. لكن هذا الشخص لا يقيس نسبة البلاد. المصطلح الصحيح:

ملاحظة مراقب الدولة

يجب أن تكون.

اعتراض 2— "إذا لم ينتج شخص ما نتيجة الدولة، لماذا نرسله؟"

لأن الملاحظة الواحدة يمكن أن:

- كشف مشاكل الوصول,
- توضيح الالتباس بين الكيانات
- حدد اللغة الخاطئة،
- ووضع الأحداث الحرجة

ظاهرة للمرة الأولى. لوحة المراقب هي نظام إنذار مبكر.

اعتراض 3 — "هل علينا مراقبة كل دولة في كل موجة؟"

لا. يمكن استخدام دورة مراقبة محددة. يمكن مراقبة الدول عالية الخطورة أو المهمة بشكل أكثر تكرارًا؛ يمكن مراقبة الدول الصغيرة ومنخفضة الخطورة بالتناوب. يجب تحديد الجدول قبل النتائج.

اعتراض 4 — "إذا كانت الدول الصغيرة تتلقى وزنًا منخفضًا في النتيجة العالمية، فلماذا نوليها كل هذا الاهتمام؟"

لأن نصيبهم من السكان قد يكون منخفضاً. ومع ذلك:

- القانون،
- اللغة،
- ضرر المستخدم المحلي،
- النزاهة الجغرافية

هي مجتمعات بشرية فعلية. لا ينبغي أن يعني الوزن المنخفض عدم الظهور.

الاعتراض 5 — "أليس إعطاء جميع اللغات نفس الوزن هو الطريقة الأعدل؟"

يمكن أن يكون الوزن المتساوي للغة منظورا تشخيصيا قيّما. إنه لا يمثل متوسط التجربة البشرية في سكان العالم. يجب نشر درجة السكان ودرجة عدالة اللغة بشكل منفصل.

الاعتراض 6 — "إذا كانت لغة صغيرة تؤثر قليلاً فقط على النتيجة العالمية، فلماذا يهم الحصول على نتيجة منخفضة؟"

لأن مشكلة التمثيل حقيقية بالنسبة للأشخاص الذين يستخدمون تلك اللغة. أيضًا، الأخطاء الشديدة في التراخيص أو الصحة أو الهوية لا تُستهان بها تلقائيًا بسبب حجم السكان.

الاعتراض 7 — "أليس من غير العادل لوم الذكاء الاصطناعي إذا كانت لغة منخفضة الموارد لديها عدد قليل من الموارد؟"

قد يكون لوم الذكاء الاصطناعي غير عادل فقط. لذلك:

- تساوي الموارد،
- تكافؤ التحفيز,
- كفاءة المحكم

تتم فحصها بشكل منفصل. الخطأ الذي يصل إلى المستخدم لا يزال صحيحًا من الناحية الواقعية. يجب تحديد مالك التصحيح بشكل صحيح.

الاعتراض 8 — "أليس من الطبيعي الحصول على إجابات مختلفة في نفس اللغة في دول مختلفة؟"

قد يكون ذلك طبيعيًا إذا كانت الظروف المحلية الفعلية مختلفة. إذا تغيرت الحقيقة المادية نفسها بطريقة مختلفة وبدون تفسير، فقد تكون هناك مشكلة. يجب فحص اللغة والمكان معًا.

الاعتراض 9 — "إذا قدم الذكاء الاصطناعي الإجابة الصحيحة بلغات أخرى، فلماذا يفشل؟"

إذا لم يستطع المستخدم فهم الرد، فلا يمكن استخدام التمثيل. يجب تقييم انحراف اللغة بشكل منفصل إلى جانب الدقة.

الاعتراض 10 — "قد يكون العثور على مستخدمي 200 من أجل تقييمات اللغة مكلفًا جدًا."

يمكن استخدام عينة أصغر. النتيجة:

- اكتشاف،
- مؤقت،
- عينة غير كافية

يجب نشره بهذه الحالة. التكلفة لا تمنح يقينًا زائفًا.

اعتراض 11— "إذا كانت النتيجة العالمية مرتفعة بالفعل، ألا تخيف بطاقة اللغة المنفصلة العميل دون داعٍ؟"

إخفاء تدهور اللغة المادي لا يحمي العميل. إنه يولد ثقة زائفة. يمكن استخدام بطاقة نتائج قصيرة وواضحة.

اعتراض 12— "هل يقلل حدث مراقب حرج واحد من جميع الأهلية؟"

يعتمد الرد التلقائي على نوع الحدث. الحدث الفردي لا يشير إلى الانتشار الوطني. ومع ذلك، إذا كان يتسبب في ضرر شديد أو سلطة زائفة، يمكن أن يؤدي إلى مراجعة نقدية واتخاذ تدابير مؤقتة. سيتم تحديد القرار في القسم 15.

القاعدة العامة للقسم 49

إنشاء نتيجة عالمية أمر سهل. تملأ سكان العالم بالدول الكبيرة. تخسر الدول الصغيرة في الكسور. تقيس اللغات ذات الموارد العالية. اللغات ذات الموارد المنخفضة: تزيلها بالقول "العينة غير كافية." ثم تنشر رقمًا واحدًا. قد يتم حساب هذا الرقم رياضيًا بشكل صحيح. لكنه قد لا يظهر كل عيوب التمثيل في العالم. NOMOS لا يعطي وزنًا متساويًا لكل دولة. لأن حصة السكان من الناس ليست متساوية. NOMOS لا يتجاهل الدول الصغيرة أيضًا. لأن قلة السكان لا تعني عدم الوجود. NOMOS لا يوزن جميع اللغات بشكل متساوٍ في النتيجة العالمية. لأن هذا قد يشوه تجربة متوسط السكان في العالم. NOMOS لا يجعل اللغات الصغيرة غير مرئية أيضًا. لأنه حتى إذا كان وزنها العالمي منخفضًا، فإن خطأ التمثيل للأشخاص الذين يستخدمون تلك اللغة يمكن أن يكون 100% حقيقيًا.

لوحة مراقبة الدول:

> تُظهر الأماكن التي لم ننظر إليها أبدًا.

لوحة عدالة اللغة:

> تُظهر من هو أقل من المتوسط العالمي.

لوحة السكان:

> تقدّر التجربة المتوسطة في سكان العالم.

عندما يعمل هؤلاء الثلاثة معًا، يصبح التقييم العالمي موثوقًا للسكان وقابلًا للتدقيق من حيث العدالة. المستخدم الفرد في بلد معين ليس هو البلد بأكمله. لكن هذا المستخدم هو شخص حقيقي. قد تكون اللغة جزءًا صغيرًا من السكان العالميين. لكن الخطأ في تلك اللغة لا يصبح أقل خطأً. قد يكون الحدث الحرج نادرًا. لكن الضرر لا يُقاس فقط بالتكرار. لذلك، فإن قانون القياس السابع لـ NOMOS هو:

> الحق في الظهور ليس هو نفسه وزن السكان.

قانونها الثامن هو كما يلي:

> مراقبة بلد ليس هو نفسه تقدير بلد.

القانون التاسع هو كما يلي:

> عدالة اللغة ليست جعل جميع اللغات متساوية؛ بل هي التأكد من أن لا أحد يتم تهميشه بشكل منهجي عن الواقع المادي بسبب لغته.

القانون العاشر هو:

> إذا كان المتوسط العالمي يجعل اللغة والبلد المتخلفين غير مرئيين، فهذا وحده ليس نتيجة جغرافية كافية.

قسم NOMOS، أمر 7

لا تخفِ البلدان التي لا تنظر إليها أبدًا أثناء عرض نتيجتي العالمية.

لا ترسل مستخدمًا واحدًا إلى بلد صغير وتجعل هذا الشخص ممثلًا للبلد بأكمله.

ولكن أيضًا لا تتجاهل بلدًا صغيرًا إلى الأبد لمجرد أن عدد سكانه قليل.

افصل بين ملاحظة البلد، وتوقعات البلد، وملاءمة البلد.

لا تحكم على البلد بأكمله بناءً على خطأ جسيم واحد. / ولكن لا تتجاهل الضرر البشري بقولك "إجابة واحدة فقط."

لا تبنِ النجاح في اللغات الكبرى على صمت اللغات الصغرى.

يمكنك أخذ عينات من لغة منخفضة الموارد كثيرًا. / لكن لا تجعلها تبدو أكبر مما هي عليه في تعداد سكان العالم.

لا تجعل درجة اللغة المتساوية درجة للسكان؛ لا تجعل درجة السكان درجة للعدالة اللغوية.

لا تلمني إذا كان الترجمة خاطئة. / إذا كان المصدر غير كامل، فقل أيضًا ذلك. / إذا لم يفهم المحكم اللغة، فلا تصدر حكمًا قاطعًا.

لا تصنّف فرقًا محليًا مشروعًا بوصفه عدم إنصاف، أو تتجاهل تباينًا ماديًا غير مفسَّر بوصفه مجرد أسلوب.

أولاً، راقب الدول. / ثم حدد الأماكن التي يمكنك فيها دعم تقدير. / ثم قم بأخذ عينات زائدة من اللغات التي تتطلب رؤية تشخيصية. / ثم أعدها إلى أوزانها الحقيقية في السكان. / اجعل الأشخاص الذين يتركهم المتوسط العالمي مرئيين.

جملة ختام الفصل

يجب أن يظهر قياس GEO عالمي حقيقي أكثر مما يراه معظم العالم. يجب أن يظهر أيضًا من يظل غير مراقب في البلدان واللغات الصغيرة، ومن يتم تمثيله بشكل خاطئ وأي نتيجة هي مجرد تنبيه واحد.

النواة المعيارية

يجب أن تظل لوحة السكان **GEO-1000**، ولوحة مراقبي الدول، ولوحة عدالة اللغة متميزة في الغرض وأسلوب الاختيار والتوزين والادعاءات العامة. يجوز لملاحظة مراقب الدولة أن تثبت حدوث شرط وصول أو استجابة أو شذوذ أو حادثة حرجة في سياق محدد لدولة–منتج–لغة. لا يجوز تمثيل ملاحظة مراقب الدولة كتقدير انتشار على مستوى الدولة أو كدرجة **GEO** للدولة. يجب الإبلاغ عن تغطية الملاحظة على مستوى الدولة، وتغطية الملاحظة الموزونة على مستوى السكان، وتغطية تقديرات الدولة بشكل منفصل. يجب أن تتلقى كل دولة مؤهلة احتمالاً إيجابياً للملاحظة ضمن دورة المراقبة المعلنة مسبقاً ما لم تكن خارج إطار المنتج القانوني نطاق الوصول الطبيعي. يجب أن تُقيّم الإنصاف اللغوي التمثيل المتقارب مادياً تحت ادعاءات متكافئة دلالياً وسياقات قابلة للمقارنة. يجب ألا يتطلب ذلك ردوداً متطابقة كلمة بكلمة أو تجاهل الاختلافات المحلية المشروعة. يجب إعادة اللغات التي تم أخذ عينات مفرطة لها إلى أوزانها في السكان المستهدفين لتقديرات السكان العالمية. يجب أن تظل الدرجات موزونة حسب السكان، والتشخيصات المتساوية للغات، وحدود اللغة، ومعدلات الأخطاء الحرجة، ومعدلات المجهول، واللغات غير المختبرة مرئية بشكل منفصل. لا يجوز عزو التفاوت اللغوي إلى منتج الذكاء الاصطناعي وحده حتى يتم تقييم تكافؤ المُدخلات، وتكافؤ المراجع، وتكافؤ القياس، وتكافؤ التحكيم. قد يحمل الملاحظة الواحدة عدة علامات لمجموعة التشخيص، ولكن لا يجوز أن تتلقى وزناً مزدوجاً للسكان بالكامل. يجوز أن تؤدي حوادث المراقب الحاسم إلى تحفيز أخذ عينات تأكيدية ومراجعة التوافق دون اعتبارها دليلاً على الانتشار في البلد.

اختيار المشاركين وترجيحهم وتحيز أخذ العينات

الأسس المنهجية / K04؛ K05؛ K06؛ K07؛ K08؛ K22

حد الفصل

الفصل 5 عرّف مجموعة المستخدمين المستهدفة. الفصل 6 حوّل هذه المجموعة إلى تصميم عينة موزع عبر الدول واللغات وأسواق المستخدمين. الفصل 7 أنشأ أنظمة مراقبة تكاملية للدول الصغيرة واللغات منخفضة الموارد التي قد يتركها اللوحة السكانية الرئيسية غير مرئية. الآن، تحتاج الأرقام في خطة العينة إلى التحول إلى أشخاص حقيقيين:

> أين سيتم العثور على هؤلاء المشاركين، وكيف سيتم اختيارهم، وتحت أي ظروف سيتم قبولهم في اللوحة، ومع أي تحيّزات سيدخلون درجة *GEO*؟

من لوحة عبر الإنترنت لمنظمة بحثية؟ من دعوة مفتوحة عبر وسائل التواصل الاجتماعي؟ من الجامعات؟ من قائمة المستخدمين الخاصة بالعميل؟ من شركاء البحث المحليين؟ فقط من الأشخاص الذين يستخدمون منتجات الذكاء الاصطناعي بشكل متكرر؟ من الذين يصلون إلى المنتج لأول مرة؟ من عملاء العلامة التجارية؟ من موظفي المنافسين؟ من مشتركي الخطط المدفوعة؟ كيف سيتم التحقق من أن الشخص يعيش فعلاً في ذلك البلد؟ كم مرة سيتم احتساب الشخص نفسه إذا كان يستخدم عدة حسابات؟ كم سيُدفع للمشارك؟ هل سيشجع الدفع إكمال المهمة بسرعة أو تقديم نتيجة محددة؟ هل سيتم الاحتفاظ بالملاحظة إذا قام المشارك بتحميل استجابة ذكاء اصطناعي غير صحيحة؟ أم سيتم استبداله بمستخدم جديد بعبارة: 'تم تنفيذ المهمة بشكل غير صحيح.'؟ قد يكون تم جمع ألف لقطة شاشة صالحة. ولكن المشاركون:

- 60% هم مستخدمون كثيفون للذكاء الاصطناعي،
- 20% هم طلاب تكنولوجيا،
- 10% هم عملاء العلامة التجارية المراجعة،
- 5% هم موظفو الشركة،
- البقية هم مستخدمون عامون

أي شريحة سكانية، إذن، سيمثلها هذا الناتج؟ هل يمكن أن تصحح الأوزان الديموغرافية الفروق بالكامل؟ مطابقة إجمالي العمر والجنس والبلد واللغة لا تصحح بالضرورة الفروق غير المقاسة في الاهتمام بالتكنولوجيا، ووعي العلامة التجارية، والوصول إلى الخطط المدفوعة أو أنماط استخدام الذكاء الاصطناعي. الأوزان مفيدة، لكنها لا تستطيع محو كل عيب في اختيار المشاركين. يطبق هذا الفصل المطلب السابق — أن يكون كل ضبط تدقيقي قابلاً للاختبار وفقًا للنموذج، التاريخ، البلد، اللغة، مجموعة الاستعلامات، عدد التكرارات وسجل القياس — على اختيار المشاركين الفعليين. كما أنه يحافظ على المطلب المركزي لـ NOMOS: الأدلة، الحدود، السياق والزمن. يتناول هذا الفصل:

- مصادر اختيار المشاركين،
- التمييز بين اللوحات الاحتمالية وغير الاحتمالية.
- التحقق من الأهلية والهوية،
- التحكم في الوصول حسب البلد واللغة والمنتج،
- تفرد المشارك،

- مخاطر المكافأة وإعادة التوجيه،
- تأثيرات لوحة التجارب،
- تعيين المهام،
- عدم الاستجابة وهدر المشاركين في اللوحة،
- المشاركة الاحتيالية أو المؤتمتة،
- التحيز في الاختيار والتغطية،
- قوة وحدود الأوزان،
- حالات جودة العينة

يحدد. هذا الفصل لم يتطرق بعد إلى:

- جميع المجالات التقنية لسجل منتجات الذكاء الاصطناعي،
- النظام الكامل للترجمة والإصدارات للصياغات مُدخل
- الهندسة المعمارية التفصيلية لبنية Capture NOMOS،
- التحكيم في الادعاء الذرية،
- صيغة النتيجة النهائية

غير محددة بالكامل. السؤال الرئيسي في القسم 8 هو:

> عندما نجمع ألف سجل صالح، كيف نثبت لأي أشخاص حقيقيين تنتمي هذه السجلات وأي عدد سكاني بشري يمكن أن تمثله؟

التحدي **NOMOS**

تقوم بشراء أشخاص 1,000 من لوحة بحث عبر الإنترنت. يخبرك مزود اللوحة: "1,000 مشارك جاهز وفقًا لحصص بلدك وعمرهم." يكمل المشاركون المهمة. يقومون بتحميل لقطات الشاشة. النتائج الخام:

> *90* في المائة تمثيل دقيق.

في تقريرك، كتبت: "90% من المستخدمين حول العالم ممثلون بدقة." ومع ذلك، فإن التوزيع الفعلي للمشاركين هو كما يلي:

تكرار استخدام الذكاء الاصطناعي	حصة السكان المستهدفين	الحصة في اللوحة
استخدام منخفض أو غير متكرر	50%	10%
استخدام متوسط	30%	30%
استخدام كثيف	20%	60%

المستخدمون الكثيفون للذكاء الاصطناعي:

- عادات أفضل في إعداد الإرشادات،
- المزيد من الحسابات المدفوعة،
- المنتجات الواجهية الأحدث،
- معرفة بالتخصيص،
- مألوفية أعلى مع التكنولوجيا

يمكن أن يكون. حتى إذا تم استخدام نفس الطلب المثبّت:

- خطة الحساب،
- واجهة
- الاستخدام السابق،
- الوصول للمنتج

قد يكون مختلفًا. دع نسب اجتياز المجموعة الاصطناعية تكون كما يلي:

تكرار استخدام الذكاء الاصطناعي	معدل النجاح
استخدام منخفض أو غير متكرر	72%
استخدام متوسط	84%
استخدام كثيف	96%

النتيجة الأولية للوحة:

```
0.10(72)+0.30(84)+0.60(96)=90
```

وهذا هو:

90 في المئة

مع أوزان السكان المستهدفين، ومع ذلك، تكون النتيجة:

```
0.50(72)+0.30(84)+0.20(96)=80.4
```

لذلك، فإن نفس ردود 1,000 تنتج نتيجتين مختلفتين:

- نتيجة اللوحة الأولية: 90 في المئة
- نتيجة السكان المستهدفين: 80.4 في المئة

تخيل الآن أنك قد وزَّنته. يحصل كل من الأشخاص 100 في مجموعة الاستخدام المنخفض على وزن مرتفع. ويحصل كل من الأشخاص 600 في مجموعة الاستخدام العالي على وزن منخفض. حجم العينة الخام لا يزال 1,000. ومع ذلك، ينخفض حجم العينة الفعّال إلى حوالي 349. لذلك، قد تكون قد جمعت 1,000 ملفات. إحصائياً، قد لا يكون لديك حجم المعلومات المكافئ للأشخاص 1,000 المتساويين والمستقلين. الآن سؤال أصعب: في مجموعة الاستخدام المنخفض، الأشخاص الذين لم يتم تضمينهم أبداً في اللوحة:

- متصلون بشكل منخفض،
- يستخدمون تكنولوجيا مساعدة،
- يعيشون في مدينة صغيرة،
- ليس لديهم جهاز محدد

إذا كان هناك أشخاص، هل يمكن أن يعيدهم الوزن؟ لا. لا يمكن إنتاج ملاحظة لشخص غير موجود في عينة الإطار على الإطلاق عن طريق الوزن. الآن ترسل الشركة المدققة اتصالات إلى عملائها: 'يرجى المشاركة في الاستبيان.' يشارك العملاء المخلصون للشركة بمعدل مرتفع. الأشخاص الذين يعرفون المنافسين لكنهم لا يستخدمون الشركة يشاركون بمعدل أقل. قد تكون أوزان العمر والبلد صحيحة. قد يستمر انحياز ولاء العلامة التجارية. الآن مشكلة أخرى: يُقال للمشارك: 'ستحصل على مكافأة إذا قمت بتحميل استجابة دقيقة وعالية الجودة من الذكاء الاصطناعي.' يقوم المشارك بمراجعة الاستجابة. يختار النتيجة الأفضل. يقتطع لقطة الشاشة. لم تعد هذه الاستجابة الطبيعية لمنتج الذكاء الاصطناعي. إنها الاستجابة المختارة التي تم إنشاؤها بواسطة نظام المكافآت. القاعدة الأولى في هذا القسم هي:

> عدد المشاركين ليس مقياسًا لقدرة تمثيل المشاركين.

ينص حكمه الثاني على:

> التعيين العشوائي ليس اختيارًا عشوائيًا للسكان.

تنص الفقرة الثالثة منه على:

> الوزن يمكن أن يقلل من الاختلال الملحوظ؛ لا يمكنه إنشاء شخص غير ملحوظ.

تنص المادة الرابعة على:

> لا يمكن أن يعتمد إدراج المشارك في اللوحة، أو احتفاظه بالمهمة، أو استبداله على محتوى استجابة الذكاء الاصطناعي.

تنص أحكامه الخامسة على:

> يجب منح المكافأة لإكمال البروتوكول بصدق، وليس لاختيار استجابة إيجابية أو مرغوبة.

1. غرض الفصل

الغرض من هذا القسم هو جعل سلسلة الاختيار بأكملها من السكان المستهدفين إلى الملاحظة الفعلية للمستخدم واضحة، غير متحيزة، وقابلة للتدقيق. يميز القسم بشكل معياري:

- السكان المستهدفون وإطار العينات
- إطار العينة مقابل مصدر المشارك
- الشخص المختار مقابل الشخص المدعو
- الشخص المدعو والشخص الذي يوافق على المشاركة
- المشارك والشخص الذي يُكمّل المهمة
- الشخص الذي يُكمّل المهمة والملاحظة الصالحة
- إنسان وحساب
- إنسان وجهاز
- إنسان وجلسة
- لوحة احتمالية ولوحة متطوعين
- الاختيار العشوائي داخل اللوحة والاختيار العشوائي من السكان
- الامتثال للحصة والتمثيل الحقيقي
- التوازن الديموغرافي والتوازن السلوكي
- الألفة الشخصية وملاءمة السكان المستهدفين
- العميل، الموظف، والمستخدم العام
- مكافأة المشاركين وحافز النتائج
- سرعة المهمة وسلامة المهمة
- ملاحظة غير صالحة مع عدم الاستجابة
- تغير السكان مع فقدان اللوحة
- مشارك مزيف وملاحظة حقيقية منخفضة الجودة
- تزوير البيانات ورد الذكاء الاصطناعي غير الصحيح
- تصحيح الانحياز باستخدام الترجيح
- إنتاج الأدلة مع المعايرة
- حجم العينة الخام مقابل حجم العينة الفعّال

- خصوصية المشاركين وأدلة التدقيق
- لوحة السكان العامة ولوحة الخبراء أو العملاء

في نهاية هذا القسم، يجب أن تكون كل مراجعة GEO-1000 قادرة على الإجابة على الأسئلة التالية:

من أي مصدر وبأي قاعدة تم اختيار المشاركين؟

من لم تتح له فرصة الانضمام إلى العينة؟

من كان أكثر احتمالًا للمشاركة ولماذا؟

ما الفروقات المعروفة التي تصححها الأوزان، وأي منها لا تصححها؟

هل تم اتخاذ قرار مشاركة غير صالح أو مزيف دون الاطلاع على محتوى رد الذكاء الاصطناعي؟

هل أثرت مكافأة المشاركين على سلوكهم واختيارهم للإجابات؟

2. الأحكام المعيارية المركزية

يجب أن تحوي كل ملاحظة GEO-1000 سلسلة اختيار مشارك قابلة للتتبع من السكان المستهدفين إلى سجل تحليلي صالح؛ يجب أن تكون قرارات الاختيار والدعوة والقبول وتكليف المهام والإكمال والتحقق والاستبعاد والتوزين مستقلة عن المحتوى الإيجابي أو السلبي لاستجابة الذكاء الاصطناعي. [K04–K06] مجلس المشاركين فقط:

- البلد،
- لا يعني أي ملاءمة من حيث العمر،
- اللغة

لا يمكن اعتباره مناسبًا لتمثيل السكان لأنه يفي بالحصص. يجب أن يحتوي المجلس أيضًا على المعلومات التالية إلى الحد الذي يكون ذا صلة:

- مصدر المشارك
- آلية الاختيار
- طريقة المشاركة في المجلس
- معايير الأهلية
- التحقق الفريد للبشر
- التحقق من الدولة واللغة
- التحقق من الوصول إلى منتج الذكاء الاصطناعي
- معلومات الحساب والخطة
- الجهاز وواجهة الاستخدام
- تكرار استخدام الذكاء الاصطناعي
- الألفة مع العلامة التجارية
- العلاقة مع الكيان الذي يتم تدقيقه
- هيكل الحوافز
- طريقة تعيين المهام
- عدم الاستجابة
- عدم الصلاحية
- التكرار وتكييف اللوحة

- الوزن
- نقص التغطية
- الخصوصية والموافقة

قد يؤثر قياس بعض هذه المعلومات على المشارك. لذلك، فإن ترتيب جمع البيانات هو أيضًا جزء من البروتوكول.

3. سلسلة اختيار المشاركين

السلسلة من السكان المستهدفين إلى الملاحظة النهائية تتكون من المراحل التالية:

U→F→S→I→E→A→C→V

هنا:

- المستخدم المستهدف
- F: الأفراد المدرجين في إطار العينة
- س: الناس المختارون
- أنا: الأشخاص المدعوون
- E: الأفراد الذين تم تأكيد أهليتهم والذين تم إدراجهم في الدراسة
- أ: الأشخاص المكلفون بالمهمة
- C: الأشخاص الذين يكملون المهمة
- V: الأشخاص الذين ينتجون ملاحظات تحليلية صالحة

في كل انتقال، قد نفقد أشخاصًا أو قد يظهر تحيز جديد. على سبيل المثال:

- أولئك في السكان المستهدفين الذين ليسوا على العينة
- أولئك على العينة الذين لم تُوجه إليهم دعوة
- أولئك المدعوين الذين لم يقبلوا
- أولئك الذين يقبلون ولكن لا يكملون المهمة
- أولئك الذين يكملون ولكن ينتجون سجلات غير صالحة

ليسوا نفس الأشخاص. لذلك، لا يمكن الإبلاغ إلا عن آخر ملاحظات صالحة 1,000. يجب أن يكون قمع الاختيار بأكمله مرئيًا.

4. مصادر المشاركين

GEO-1000 لا يجب أن يكون مرتبطًا بمصدر مشارك واحد فقط. كل مصدر يحمل تحيزه ونطاقه الخاص.

4.1. لوحة احتمال السكان

للمشاركين احتمال معروف وغير صفري ليتم اختيارهم من السكان المحددين. المصادر:

- الأسرة،
- العنوان،
- السجل السكاني أو المؤهل،
- احتمال الهاتف أو اختيار متعدد المراحل

قد يكون. المزايا:

- يمكن تأسيس وزن التصميم،
- الاستنتاج السكاني أقوى،

يمكن أن تكون عملية الاختيار أكثر شفافية. القيود:

- يمكن أن يكون مكلفًا على المستوى العالمي،
- في بعض الدول، قد لا يكون هناك إطار مناسب،
- قد يكون من الضروري أيضًا العثور على الشريحة السكانية التي لديها وصول إلى منتجات الذكاء الاصطناعي،

مرة أخرى، سيكون هناك عدم استجابة. الاختيار الاحتمالي ليس دائمًا تمثيلًا مثاليًا.

4.2. لوحة البحث التطوعي المهنية

لقد سجل المشاركون مسبقًا طوعًا في لوحة البحث. المزايا:

- سريع،
- يمكن الوصول إلى عدة دول،
- يمكن تقديم الملف الشخصي والمعرفة باللغة،

يمكن إجراء القياس المتكرر. القيود:

- المشاركون معرضون للبحث،
- قد يكون مستوى استخدام التكنولوجيا مرتفعًا،
- قد يكون هناك مستخدمون كثيفون للوحة،
- قد تكون احتمالات الدخول إلى المجتمع غير معروفة،

قد يكون نفس الشخص مسجلاً في أكثر من لوحة. هذه اللوحة:

لا يمكن تقديمها كعينة سكان عالمية احتمالية

يجب ألا يتم تمثيلها كعينة احتمالية للسكان العالميين.

4.3. دعوة مفتوحة على الإنترنت

يتم الحصول على المشاركة من خلال وسائل التواصل الاجتماعي أو موقع إلكتروني أو نموذج تسجيل مفتوح. المزايا:

- الرؤية الواسعة,
- منخفض التكلفة,
- الوصول إلى الدول الصغيرة ومجتمعات اللغة

يمكن توفيره. القيود:

- اختيار ذاتي قوي،
- تركيز على معجبي أو نقاد العلامة التجارية،
- المشاركة المزيفة والمتكررة،
- التمثيل الزائد لعشاق التكنولوجيا

يمكن أن يتكوَّن. قد يكون الدعوة المفتوحة أكثر ملاءمة بكثير من لوحة السكان الرئيسية لـ:

- اكتشاف،
- مراقب الدولة،
- العدالة اللغوية،

- تحدي التكرار

قد تكون هذه المصادر أكثر ملاءمة لتلك الأغراض.

4.4. لوحة جامعة أو مؤسسة مرتبطة

المشاركون:

- يتم اختيارهم من خلال الجامعات,
- مؤسسة عامة،
- مراكز البحوث،
- المنظمات المهنية

المزايا:

- الوصول المحلي
- الخبرة اللغوية
- الرقابة الأخلاقية
- الوصول إلى المجتمعات منخفضة الموارد

يمكن توفيره. القيود:

- شدة الطالب أو المحترف,
- تأثير المؤسسة
- التحيّز بناءً على العمر والتعليم
- ضغط المشاركة المؤسسية

يمكن إنشاء.

4.5. قائمة العملاء أو المستخدمين

يتم اختيار العملاء الفعليين أو المستخدمين للمؤسسة التي تم تدقيقها. هذا المصدر:

- تجربة العملاء،
- ملاءمة للخدمة،
- الوعي بالعلامة التجارية

إنه ذو قيمة للقياس. إنه ليس لوحة عامة للسكان. العملاء:

- يمكنهم التعرف على العلامة التجارية بشكل أفضل،
- قد تكون لديهم تجربة أكثر إيجابية أو أكثر سلبية،

قد يكونون أكثر احتمالًا لطرح الاستفسار ذي الصلة. يجب تعريف لوحة العملاء بشكل منفصل.

4.6. المشاركون المقترحون من المؤسسة التي تم تدقيقها

الشركة:

- الموظف،
- العميل،
- الشريك،
- متابع

يمكن الاقتراح بها. لا يمكن تضمين هؤلاء المستخدمين في لوحة السكان المستقلة الرئيسية بدون تفسير. بشكل منفصل:

عينة مرشحة من الكيان

يجب تصنيفها. إذا اختارت المؤسسة المشاركين، فإن النتيجة ليست قياسًا مستقلًا للسكان.

4.7. عينة هجينة متعددة المصادر

إنها مزيج من مصادر متعددة. مثال:

- عينة وطنية قائمة على الاحتمالات،
- عينة مهنية على الإنترنت،
- شريك محلي في بلد صغير،
- مجتمع عدالة اللغة

يمكن استخدامها معًا. يمكن للّجنة الهجينة زيادة التغطية. ومع ذلك، من بين الموارد:

- الأفراد المتكررين،
- احتمالات الاختيار،
- الحوافز,
- خبرة المهمة،
- ملف المستخدم

يجب أيضًا نمذجته.

5. حالات نزاهة اختيار المشاركين

يمكن تصنيف جودة اختيار المشاركين في لوحة وفقًا للحالات التالية.

PSI-0— اختيار غير معروف

غير معروف بما فيه الكفاية أين وكيف تم اختيار المشاركين. لا يمكن إجراء تعميم على السكان.

PSI-1— اختيار حر أو مفتوح

إنه عينة من الراحة أو التطوع المفتوح. يمكن استخدامها للاستكشاف واكتشاف الأحداث.

PSI-2— اختيار حر تحت السيطرة الحصصية

هناك حصص تعتمد على البلد أو اللغة أو الديموغرافيا. احتمال الدخول إلى المجتمع غير معروف.

PSI-3— لوحة متعددة المصادر معايرة

هناك مصادر متعددة، حصص محددة مسبقًا، ومعايرة للسكان. قد يستمر وجود تحيز اختيار غير مقاس.

PSI-4— احتمالية أو تصميم هجين قوي

هناك احتمال معروف للاختيار أو تصميم هجين قوي يجمع بين الاحتمالية وغير الاحتمالية.

PSI-5— اختيار مكرر وتم تدقيقه خارجيًا

تصميم الاختيار:

- مع مزودين مختلفين،
- في موجات مختلفة،
- تحت تدقيق مستقل

تم فحص السلوك المتكرر والمتحيز. هذه الحالات لا تضمن دقة النتيجة. إنها تشير إلى مستوى الثقة الناتج عن اختيار المشاركين.

6. اختيار عشوائي ضمن المجموعة واختيار عشوائي من السكان

قد يكون هناك 100,000 مستخدم مسجل في لوحة المتطوعين. من هذه اللوحة، يمكن اختيار 1,000 شخصًا بشكل عشوائي. الاختيار:

هو عشوائي ضمن اللوحة.

ومع ذلك، فإن اللوحة نفسها لم تُشكّل بشكل عشوائي من سكان العالم. لذلك، لا يمكن القول: "1,000 شخصًا بشكل عشوائي من سكان العالم." التعبير الصحيح هو:

"يتم اختيار المشاركين *1,000* عشوائيًا ضمن لوحة متطوعين مسجلة مسبقًا وتتم معايرتهم لتناسب الفئة المستهدفة."

يجب أن تقتصر العشوائية على المرحلة التي يتم تطبيقها فيها.

7. سجل أهلية المشاركين

يجب أن يستوفي كل مشارك الشروط التالية إلى الحد المناسب قبل أن يتم تضمينه في اللوحة الرئيسية:

- أهلية العمر
- موافقة المتطوع
- ملاءمة البلد والمكان
- إجادة اللغة
- الوصول إلى منتج الذكاء الاصطناعي
- الخطة الصحيحة أو واجهة المستخدم
- الجهاز الضروري والوصول التقني
- حالة إنسانية فريدة
- استقلالية المهمة
- إعلان المصالح أو العلاقات
- أمن البحوث
- الموافقة على جمع البيانات والأدلة

لا يمكن تغيير قرار المطابقة بأثر رجعي بعد توليد استجابة الذكاء الاصطناعي.

8. التحقق من الدولة

السؤال "إلى أي دولة ينتمي المشارِك؟" ليس له إجابة واحدة. يجب فصل الحقول التالية:

- دولة الإقامة المعتادة
- الدولة الفعلية وقت القياس
- البلد أو منطقة السوق لحساب الذكاء الاصطناعي
- دولة الدفع أو الخطة
- المواطنة

- البلد المستهدف في صياغة المُدخل
- الموقع

قد لا تكون المواطنة هي المتغير الأساسي لمعظم القياسات الجغرافية. ما هو مهم في معظم الأحيان:

هو الحالة الجغرافية والمنتج الفعلية التي يختبر فيها المستخدم المنتج.

8.1. المستخدم المسافر

قد يكون المشارك في دولة غير دولته المعتادة. هذا الشخص:

- السكان المقيمون،
- موقع القياس،
- المستخدم المسافر

يجب أن يحتفظ بحالة منفصلة كـ.

8.2. VPN أو مغير الموقع

استخدام VPN:

- الخصوصية،
- شبكة المؤسسة،
- عادة المستخدم العادي

قد يكون السبب. لا ينبغي اعتباره احتيالًا تلقائيًا. ومع ذلك، فإنه يصعب التحقق من طبقة البلد. قد توجد قاعدة في لوحة التحكم الرئيسية للبلد للتحكم في إيقاف أو تسجيل استخدام VPN. تغييرات الموقع المستخدمة لتجاوز القيود غير مناسبة للوحة نطاق الوصول الطبيعي الرئيسية.

8.3. مستويات التحقق من البلد

GV-0— غير موثق

يتوفر فقط الإقرار الذاتي.

GV-1— تم التحقق من قبل اللوحة

مزود اللوحة لديه تسجيل بلد موجود.

GV-2— متسق مع الجلسة

الإعلان والموقع وإشارات الجلسة متسقة.

GV-3— تم التحقق من الإشارات المتعددة

متوافق مع حماية الخصوصية ويعمل مع إشارات متعددة.

GV-4— تم تدقيقه بشكل مستقل

طريقة التحقق من البلد قد اجتازت الرقابة الخارجية للجودة. جمع عنوان شارع دقيق أو إحداثيات حساسة ليس مطلبًا افتراضيًا. يجب إجراء التحقق من البلد مع تقليل البيانات.

9. إتقان اللغة والاستخدام الطبيعي

قد يعتمد معرفة المشارك باللغة فقط على العبارة: "نعم، أنا أعرف." يمكن أن يكون هذا تسجيلًا أوليًا منخفض المخاطر. في عينة لغة عالية الأمان، بالإضافة إلى ذلك:

- بيان الاستخدام الطبيعي،
- فحص فهم مختصر،
- تاريخ العينة،
- تقييم الحكم المحلي،
- تجربة محايدة قبل المهمة

قد يتم استخدام ذلك. لا ينبغي أن يتحول اختبار اللغة إلى امتحان أكاديمي ثقيل يختار أشخاصًا متعلمين تعليماً عالياً فقط. الهدف:

> لكي يفهم المشارك صياغة المُدخل والاستجابة كمستخدم حقيقي.

10. التحقق من الوصول إلى منتج الذكاء الاصطناعي

المشارك:

- يجب أن يكون قادرًا على الوصول إلى منتج الذكاء الاصطناعي الصحيح،
- الـ واجهة المستخدم الصحيح،
- خطة الحساب أو حالة الحساب الصحيحة

يجب أن يكون قادرًا على الوصول إليها. يجب تسجيل الحالات التالية بشكل منفصل:

- الخطة المجانية
- خطة فردية مدفوعة
- خطة المؤسسات
- خطة التدريب
- الويب
- الهاتف المحمول
- سطح المكتب
- المنتج المدمج
- النموذج المرئي
- ميزة الويب/الاسترجاع

لا يمكن للمشارك استخدام حساب شخص آخر. لا يمكن طلب مشاركة كلمة المرور. يجب التحقق من إثبات الوصول إلى المنتج دون الكشف علنًا عن هوية الحساب.

11. العلاقة بين المشارك والكيان الذي يتم تدقيقه

قد تؤثر علاقة المشارك بالمنظمة المراقبة على سلوك التقاط الاستجابات أو التعرف على العلامة التجارية. حالات العلاقة:

الجمهور العام

العميل الحالي

العميل السابق

العميل المحتمل

الموظف

الموظف السابق

المقاول

الشريك

المستثمر

شركة تابعة لمنافس

مزود الذكاء الاصطناعي التابع

باحث

مجهول

العلاقات المباشرة بين الموظفين وصناع القرار في لوحة السكان العامة الرئيسة:

- يمكن استبعادها،
- يمكن وضعها في طبقة منفصلة،

يمكن حمايتها بوزن منخفض من السكان الفعليين. قد يكون استبعاد جميع الأفراد ذوي الصلة تلقائيًا خطأ أيضًا. إذا كانوا جزءًا من السكان الفعليين للعملاء أو الموظفين، يمكن قياسهم بشكل منفصل وصريح.

12. معرفة العلامة التجارية

الجهة المشاركة المدققة:

- لم تسمع بها من قبل،
- يعرف الاسم،
- يعرف المنتج،
- هو زبون،
- قد يكون لديه رأي إيجابي أو سلبي قوي.

هذه المعلومات:

- يمكن أن يؤثر على سلوك المشارك,
- عينة المستخدمين في ظروف الاستخدام الطبيعية
- استجابة سريعة

المطلوب. إذا تم قياس الألفة قبل الإشارة، فقد يهيئ المستخدم للعلامة التجارية. لذلك، ترتيب التفضيل:

- معلومات اللوحة الموجودة سابقًا
- سؤال منفصل بعد المهمة
- إذا لزم الأمر، قياس مسبق محايد لا يوجه العلامة التجارية

يمكن إجراؤها. يجب أن تكون الألفة واضحة في العينة الرئيسية؛ يجب ألا يتم اختيارها كمتغير للوزن بعد رؤية النتائج.

13. ألفة استخدام الذكاء الاصطناعي

المشاركون:

- مستخدم لأول مرة،
- مستخدم نادر،
- مستخدم منتظم،
- مستخدم مكثف أو محترف

يمكن أن يكون الأفراد. تواتر استخدام الذكاء الاصطناعي:

- الوصول إلى المنتج،
- نوع الخطة،
- معرفة الواجهة،
- التخصيص،
- احتمالية إتمام المهمة بشكل صحيح

يمكن أن يتأثر. ومع ذلك، فإن اختيار مستخدمي الذكاء الاصطناعي ذوي الخبرة فقط يمكن أن يقلل من أخطاء البروتوكول. لكن ذلك قد يشوه تقدير السكان الطبيعي. الحل الصحيح:

- قياس مستوى الاستخدام،
- مقارنته مع السكان المستهدفين،
- تصنيف أو وزن إذا لزم الأمر،

والحفاظ على الخبرة الواقعية الصالحة للمستخدم غير الخبير.

14. قاعدة بشرية فريدة

وحدة السكان لـ GEO-1000 هي إنسان. نفس الشخص:

- في عدة لوحات،
- مع عدة عناوين بريد إلكتروني،
- مع عدة حسابات ذكاء اصطناعي،
- من أجهزة مختلفة

يمكنه المشاركة. يجب إجراء فحوصات التفرد بطريقة تحافظ على الخصوصية. الإشارات التي يمكن استخدامها:

- هوية المشارك المجهول
- إعادة التحقق باستخدام التشفير أو التجزئة
- متوافقية الجهاز والجلسة
- سجل الدفع أو الحوافز
- نمط الوقت والمهام

- نفس لقطة الشاشة أو ملف الاستجابة
- إعادة التحقق الآمنة عبر مزودي لوحات متعددة

لا تحتاج أي إشارة واحدة إلى أن تكون كافية للاستبعاد التلقائي. على وجه الخصوص:

- الأسر التي تشارك جهازًا مشتركًا,
- شبكة المؤسسة،
- أدوات الوصول
- مستخدمو الأجهزة المشتركة من ذوي الدخل المنخفض

قد يُعتبر خطأً مزيفًا. يجب أن تخضع السجلات المشبوهة لمراجعة بشرية وعملية استئناف.

15. مكافأة المشارك

قد يؤثر الدفع أو المكافأة التي تُمنح للمشارك على سلوك بياناته.

15.1. الدفع مستقل عن النتيجة

الطريقة المفضلة:

```
دفع ثابت ومعلن مسبقًا لإتمام المهمة وفقًا للبروتوكول.
```

الدفع:

- لا يمكن الاعتماد على أن يكون الجواب إيجابياً،
- أن يُجبر على الانتماء إلى فئات

صحيح،

- طويل،
- مُؤَيّد للعلامة التجارية
- أو مصحوبة بالاستشهادات.

يجب أن يظل الاختيار مستقلاً عن جميع هذه النتائج.

15.2. مكافأة الجودة

إذا كان من المزمع استخدام مكافأة الجودة، فقط:

- الأدلة الكاملة،
- صياغة المُدخل الصحيحة،
- المنتج الصحيح،
- التسليم في الوقت المحدد،
- تنظيف البيانات الشخصية

يمكن ربطها فقط بشروط البروتوكول. لا يمكن ربطها بمحتوى رد الذكاء الاصطناعي.

15.3. حافز السرعة

المكافأة فقط للسرعة:

- السرعة،
- الالتقاط غير الكامل

- المنتج الخاطئ،
- قطع الإجابة

قد يزيد من المخاطر. يجب السماح بوقت معقول ضمن نافذة الميدان.

15.4. الدفع حسب البلد

قد يختلف الدفع وفقًا للرسوم المحلية وظروف المعيشة. هذا الوضع:

- قد يؤثر على احتمالية المشاركة،
- تكوين العينة

ويجب تسجيله مع نهج الدفع والفروق بين البلدان.

15.5. الإكراه والحافز المفرط

يجب ألا يخلق المكافأة ضغطًا لا يستطيع الشخص رفضه بشكل معقول. على وجه الخصوص:

- الدخل المنخفض،
- طالب,
- الموظفون التابعون،
- الفئات الضعيفة

يجب حماية طوعية مشاركتهم.

16. نص الدعوة والتوجيه المسبق

عند دعوة مشارك، قد يُعد غرض البحث لهم نتيجة معينة. دعوة خاطئة: "نحن نختبر كيف تفسر أنظمة الذكاء الاصطناعي شركة Apple بشكل غير صحيح." قد يؤدي هذا البيان إلى أن يبحث المشارك عن الأخطاء. دعوة خاطئة أخرى: "ساعدنا في إثبات نجاح الذكاء الاصطناعي لعلامتنا التجارية." يخلق هذا البيان توقعًا لنتيجة إيجابية. الدعوة المحايدة المفضلة: "دراسة تجربة المستخدم الموحدة وملاحظة التمثيل في منتجات ذكاء اصطناعي محددة." قد يصبح الكيان الذي يتم تدقيقه مرئيًا لاحقًا بسبب طبيعة المهمة. ومع ذلك، لا ينبغي الكشف عن النتيجة المرغوبة.

17. تكليف المهمة

المشاركون:

- منتج الذكاء الاصطناعي،
- مُدخل,
- اللغة،
- موجة,
- ظرف الجلسة،
- ترتيب المهام

يجب تعيينها مسبقًا باستخدام طريقة محددة مسبقًا. للمشارك:

- منتجات الذكاء الاصطناعي المفضلة لديه،
- الخطة التي يعرفها جيدًا،
- اللغة التي لديها احتمال عالٍ للحصول على استجابة إيجابية

إذا سُمِح لهم بالاختيار بأنفسهم، تحيز التعيين قد يحدث.

17.1. التعيين العشوائي داخل الطبقة

بعد التحقق من الدولة واللغة والوصول إلى المنتج، يمكن تعيين المهام عشوائيًا ضمن الطبقة.

17.2. التعيين على شكل كتل

يمكن استخدام الكتل لموازنة الخطط المجانية/المدفوعة، أو التصفح عبر الهاتف/الويب، أو تكرار استخدام الذكاء الاصطناعي.

17.3. ترتيب المنتج

إذا كان الشخص يختبر أكثر من منتج، يجب أن يكون ترتيب المنتجات عشوائيًا. قد تؤثر استجابة المنتج الأول على تقييم الاستجابة التالية أو سلوك المهمة.

17.4. ترتيب صياغات المُدخل

إذا تلقى المشارك أكثر من مطالبة:

- التعلم،
- اكتساب معلومات جديدة عن العلامة التجارية،
- حمل الإجابة السابقة إلى السؤال التالي

ينشأ خطر. يجب تسجيل ترتيب صياغات المُدخل وموازنته إذا لزم الأمر.

18. تكييف العينة

المستخدمون الذين يشاركون بشكل متكرر في لوحات البحث قد يتعلمون توقعات المهمة. الشخص الذي يشارك مرارًا في نفس الاختبار:

- قد يتذكر الإجابة الصحيحة,
- قد تختار لقطة الشاشة بعناية أكبر،
- قد يراجع موقع الجهة التي تمت مراجعتها مسبقًا,

يمكن معرفة أي الأخطاء يتم البحث عنها. يطلق على هذا الوضع:

تهيئة اللوحة

يمكن أن يغير تهيئة اللوحة تجربة المستخدم الطبيعية الحقيقية.

18.1. مصادر التكييف

الموجات السابقة لـ GEO إعادة قياس الكيان نفسه بنك الأسئلة نفسه تدريب مراجعة النظراء مشاركة المهام في مجتمعات المشاركين المدفوعات السابقة وملاحظات الجودة

18.2. ضوابط التكييف

حالة المشارك الجديد والمتكرر تاريخ المشاركة الأول التعرض السابق لنفس الكيان التعرض السابق لعائلة الأسئلة نفسها لوحة دوارة عشوائية ترتيب المهام فترة راحة محددة عينة فرعية من المستخدمين الجدد مقارنة نتائج المستخدمين المشروطين والجدد إعادة استخدام نفس المشارك غير محظورة. إنها ذات قيمة للقياس الطولي. لا يمكن تقديمها كلوحة مستعرضة طبيعية.

19. تلوث المهام واتصال المشاركين

المشاركون:

- صياغات المُدخل,
- الاستجابات،
- اللقطات المصورة،
- النتائج المتوقعة

يمكن مشاركتها مع بعضهم البعض. قد تؤدي هذه الحالة إلى تعطيل الملاحظات المستقلة. الضوابط:

- تحديد مدة الموجة
- توزيع المهام بشكل عشوائي
- الالتزام بعدم مشاركة النتائج مع المشاركين
- البحث عن الإجابات والصور المكررة
- فحص النصوص غير المعتادة والتشابهات الزمنية
- مراقبة تحذيرات المجتمع أو مقدم اللوحة

التواصل مع المشاركين ليس احتيالاً بشكل تلقائي. يجب فحص تأثيره المادي.

20. أنواع عدم الاستجابة

عدم الاستجابة ليس حدثًا واحدًا.

20.1. عدم الرد على الدعوات

المشارك لا يرد على الدعوة للبحث على الإطلاق.

20.2. رفض المشاركة

الشخص يرى الدعوة ويختار عدم المشاركة.

20.3. عدم بدء المهمة بعد التأهيل

تم العثور على الشخص مناسبًا لكنه لم يبدأ المهمة.

20.4. انقطاع المهمة

المشارك يترك المهمة في منتصف الطريق.

20.5. عدم تحميل الأدلة

قد يتم الحصول على استجابة من الذكاء الاصطناعي، ولكن السجل المطلوب لم يتم رفعه.

20.6. إتمام غير صالح

تم إكمال المهمة، ولكن:

- طلب خاطئ،
- المنتج الخاطئ،
- إجابة مختارة،
- صورة مفقودة

هي أسباب لعدم الصلاحية. يجب الإبلاغ عن نسب هذه المراحل بشكل منفصل.

21. قمع الاستجابة

للمستوى h:

- nhsel: الشخص المختار
- nhinv: الشخص المدعو
- nhcon: الشخص الذي وافق على المشاركة
- nhsta: الشخص الذي بدأ المهمة
- nhcom: الشخص الذي أتم
- nhval: ملاحظة صالحة

لنكن. المعدلات الأساسية:

```
RR_h^invite= n_h^con/n_h^inv; RR_h^complete= n_h^com/n_h^sta; RR_h^valid= n_h^val/n_h^com
```

يمكن توضيحها كما يلي. كل من هذه المعدلات يشير إلى مشكلة مختلفة. إعطاء عدد القيم النهائي الصالح فقط يخفي انحياز المجال.

22. انحياز عدم الاستجابة

قد يختلف الأشخاص الذين لا يستجيبون عن أولئك الذين يستجيبون. مثال: المستخدم ذو الإنترنت الضعيف لا يمكنه إكمال مهمة. مستخدم الخطة المجانية لا يمكنه الوصول إلى المنتج. المستخدم غير المتمرس يجد التعليمات صعبة. المستخدم الكثيف لاستخدام الذكاء الاصطناعي يكمل المهام بشكل أسرع. في بعض البلدان، يبقى الدفع منخفضًا. المستخدم الذي لديه تجربة سلبية مع العلامة التجارية قد يكون أكثر استعدادًا للمشاركة. حتى لو كانت نسبة عدم الاستجابة عالية، قد يكون الانحياز منخفضًا. حتى لو كانت نسبة عدم الاستجابة منخفضة، فإن مجموعة صغيرة مفقودة قد تكون مهمة ماليًا. لذلك، ليس فقط المعدل وحده:

ما الخصائص التي تختلف بين المستجيبين وغير المستجيبين

يجب فحصها.

23. أوزان عدم الاستجابة

إذا تم استخدام إعداد عدم الاستجابة، يجب تحديد الفئات قبل مشاهدة النتائج. المتغيرات المحتملة:

- البلد
- اللغة
- العمر
- تكرار استخدام الذكاء الاصطناعي
- الخطة
- واجهة
- مصدر اللوحة
- إمكانية الوصول
- الألفة مع العلامة التجارية

معامل عدم الاستجابة:

```
a_r^{NR}= (Σ_{j∈S_r} d_j)/ (Σ_{i∈R_r} d_i)
```

يمكن حسابه على هذا النحو. ومع ذلك، فإن هذا التعديل يعتمد فقط على الافتراض داخل نفس الفئة: أن المشاركين يمثلون بما فيه الكفاية غير المشاركين. ولا يحل الاختلافات غير المقاسة.

24. آليات البيانات المفقودة

يمكن اعتبار الغياب من الناحية المفاهيمية ضمن ثلاث فئات.

24.1. فقدان عشوائي

الغياب ليس مرتبطًا جوهريًا بالنتيجة وخصائص المستخدمين. في الواقع، قد يكون نادرًا.

24.2. الغياب مفسر بالميزات الملاحظة

النقص:

- البلد،
- لا يعني أي ملاءمة من حيث العمر،
- الخطة،
- اللغة

مرتبط بمناطق معروفة جيدًا مثل. يمكن أن يساعد الترجيح جزئيًا.

24.3. النقص مرتبط بالميزات غير الملاحظة

على سبيل المثال:

- عدم الثقة في منتج الذكاء الاصطناعي،
- ضعف المهارات الرقمية،
- معارضة العلامة التجارية،
- القلق من الخصوصية

إذا لم يتم قياسها، لا يمكن أن يصحح الترجيح النقص بالكامل. في هذه الحالة، هناك حاجة إلى تحليل الحساسية.

25. فقدان المشاركين وتباعد المراحل الزمنية

في اللوحات الطولية أو الدوارة حيث يتم قياس نفس المستخدمين مرة أخرى، بعض الأفراد لا يشاركون في الموجة التالية. هذا هو فقدان المشاركين. أولئك الذين يبقون في اللوحة:

- أكثر انضباطًا،
- أكثر دافعًا للمكافأة،
- أكثر خبرة،
- أكثر تعودًا على البحث

قد يكونون. يمكن أن ينشأ التغير في النتيجة مع مرور الوقت جزئيًا من تركيب اللوحة.

25.1. سجل فقدان المشاركين في اللوحة

في كل موجة:

- تمت دعوته مرة أخرى،
- شارك،
- انسحب،
- تمت الإضافة حديثًا

يجب عرض أرقام المستخدمين.

25.2. وزن تسرب اللوحة

يمكن وزن المشاركين المتبقين وفقًا لاحتمالية الانسحاب. هذه الطريقة لا تحل تمامًا التحيز الناتج عن تسرب اللوحة غير المرصود.

25.3. تحديث اللوحة

يمكن إضافة مستخدمين جدد لاستبدال المستخدمين المفقودين. يجب الإبلاغ عن نتائج المستخدمين الجدد والعائدين بشكل منفصل.

26. احتيال ومشاركة المشاركين الآلية

هناك خطر من الاحتيال عندما تكون GEO-1000 لوحة محفزة وعالمية. السلوكيات المحتملة:

- شخص يشارك بحسابات متعددة
- استخدام الروبوتات التلقائية
- تحميل لقطة شاشة لشخص آخر
- تعديل الصور
- تحميل نفس الإجابة نيابةً عن مستخدمين مختلفين
- الإعلان عن البلد أو اللغة بشكل خاطئ
- مشاركة الحساب
- جعل شخص آخر يقوم بإكمال المهمة
- إعادة الانضمام بين مزودي اللوحة
- إنشاء مزارع مهام جماعية

هذا الخطر لا يسمح باعتبار جميع المشاركين منخفضي الدخل أو السريعين مشبوهين. اكتشاف الاحتيال:

- المناظرات،
- إشارات متعددة،
- مراجعة بشرية نسبية

يجب أن يكون على أساس.

27. حالات الاحتيال والنزاهة

FI-0— لم يتم التقييم

لم يتم تقييم السجل من حيث النزاهة.

FI-1— لا توجد شذوذات مادية

لم يتم العثور على علامات واضحة للاحتيال. هذه ليست ضمانة مطلقة للدقة.

FI-2— يتطلب مراجعة

هناك إشارة أو أكثر مريبة.

FI-3— مشكوك في التكرار أو التواطؤ

هناك شك في تكرار الشخص أو البصر أو الاستجابة أو التنسيق.

FI-4— تم تأكيد التلاعب بالبروتوكول

لقد قام المشارك بتغيير البروتوكول عن عمد.

FI-5— تم تأكيد الملاحظة الملفقة

الملاحظة ليست تجربة مستخدم ذكاء اصطناعي حقيقية. FI-2 ليست استبعادًا تلقائيًا. المراجعة مطلوبة.

28. إشارات احتيال

نفس تجزئة لقطة الشاشة
نفس بصمة الملف النادر
وقت إتمام غير ممكن
تكرار هويات متعددة من نفس الجهاز أو الجلسة
تضارب في المادة حسب الدولة، المنطقة، ووصول المنتج
عدم تطابق بين نص الاستجابة والصورة
أثر تعديل الصورة
نفس الشخص في القائمة الأساسية والاحتياطية
نفس الشخص المجهول عبر عدة مزودين للوحة
إتمام المهام بطريقة متسلسلة وميكانيكية
نفس نمط الخطأ غير المعتاد
قد لا تكون إشارة واحدة دليلًا قاطعًا على الاحتيال.

29. العدالة في اكتشاف الاحتيال

قد يتم تمييز المستخدمين التاليين عن طريق الخطأ كأشخاص محتالين:

- العائلات التي تشارك جهازًا واحدًا
- الموظفون الذين يستخدمون شبكة مؤسسية
- مشاركو مختبر الجامعة
- الأفراد الذين يحرصون على الخصوصية ويستخدمون شبكات VPN
- المستخدمون من ذوي الإعاقة الذين يستخدمون برامج قراءة الشاشة أو مساعدات الأتمتة
- الذين يرفعون صورًا مضغوطة متشابهة بسبب ضعف عرض النطاق

لذلك:

- tagging التلقائي,
- المراجعة البشرية,
- وصف المشارك إذا لزم الأمر,
- الاستئناف

يجب أن يعملوا معًا.

30. يجب فصل صحة المشارك عن دقة رد الذكاء الاصطناعي

قد يكون سجل المشارك متوافقًا مع البروتوكول. قد يكون رد الذكاء الاصطناعي غير صحيح. هذه ملاحظة صالحة.
قد يكون رد الذكاء الاصطناعي لمشارك آخر صحيحًا. ومع ذلك، المستخدم:

- قام بتحديث الإجابة ثلاث مرات،
- قام بتحميل أفضل إجابة فقط،
- قد يكون قد غيّر صياغة المُدخل

هذه ملاحظة غير صالحة. لذلك:

> الإجابة الصحيحة للذكاء الاصطناعي لا تعني أن البيانات صحيحة. / الإجابة غير الصحيحة للذكاء الاصطناعي لا تعني أن البيانات غير صالحة.

يجب إجراء تقييم الصلاحية قبل التقييم الموضوعي وبأقصى قدر ممكن من الحياد.

31. تضارب المصالح ومصدر المشارك

قد يستفيد المشارك أو مزود اللوحة من النتائج التالية:

- الشركة التي تم تدقيقها تحصل على درجة عالية
- منتج الذكاء الاصطناعي يظهر متفوقًا
- المنافس يحصل على درجة منخفضة
- بيع المزيد من مهام التدقيق
- جائزة أو علاقة عمل

لذلك:

- علاقات المشارك
- اتفاقية مزود اللوحة
- الأجر المستقل عن النتائج
- تدخل العميل

يجب شرحه. المؤسسة التي خضعت للتدقيق:

- أي المستخدمين يجب دعوتهم،
- من سيتم اعتباره غير صالح،
- من سيتم استبداله باحتياطي

لا يمكن اتخاذ القرار.

32. تأثير مزود اللوحة

قد يكون مستخدمو مزودي لوحات اثنين في نفس البلد مختلفين. مزود واحد:

- محبو التكنولوجيا،
- الطلاب،

- المشاركون في الأبحاث المكثفة

قد يكون لديهم معدل أعلى من حيث العناية. يجب الاحتفاظ بمزود اللوحة كمتغير عنقودي أو متغير مصدر في التحليل. إذا أمكن:

- النتائج المستندة إلى المزود,
- تأثير المصدر،
- مقارنة المزودين داخل نفس البلد

يجب فحصها. لا يمكن تقديم لوحة عالمية بمزود واحد على أنها سكان العالم أنفسهم.

33. تكرار الفرد في لوحة متعددة المصدر

نفس الفرد:

- لوحتان مهنيتان،
- في دعوة مفتوحة،
- في قائمة الجامعة

يمكن العثور عليها. يمكن للوحة متعددة المصادر زيادة عدد السجلات دون زيادة التفرد. التحقق من التكرار عبر المصادر:

- مع الحفاظ على الخصوصية،
- دون الحاجة إلى الإفصاح عن الهوية المركزية.
- السماح بالاعتراض على عدم التطابق

يجب تأسيسه بطريقة.

34. خريطة تحيز العينة

ينبغي لمنهج GEO-1000 رصد أنواع التحيز الآتية كلٌّ على حدة.

SB-1— تحيز التغطية

بعض شرائح السكان المستهدفة غير موجودة في إطار العينة.

SB-2— تحيز الاختيار

احتمالية اختيار الأفراد في الإطار للمشاركة في اللوحة لا تتوافق مع التصميم المقصود.

SB-3— تحيز الاختيار الذاتي

المتطوعون للدراسة يختلفون بشكل منهجي عن السكان.

SB-4— تحيز الدعوة والإرشاد

نص الدعوة أو المصدر يجذب الأشخاص ذوي وجهة نظر محددة.

SB-5— تحيز عدم الاستجابة

الأشخاص الذين لا يشاركون أو لا يكملون يختلفون من حيث النتائج.

SB-6— تحيز تراجع أعضاء اللوحة

المستخدمون الذين يبقون بين الجولات يختلفون عن أولئك الذين ينسحبون.

SB-7— تحيز الحافز

الدفع أو المكافأة تؤثر على سرعة المشارك أو سلوك الاختيار أو الاستجابة.

SB-8— تكييف اللوحة

المشارك يتعلم السلوك المتوقع من الدراسات السابقة.

SB-9— تحيز تعيين المهمة

يختار المشاركون المنتج أو اللغة أو التعليمات بأنفسهم أو يتم تعيينهم بشكل غير متساوٍ.

SB-10— تحيز تصنيف البلاد واللغة

يتم تعيين المشارك إلى خلية بلد أو لغة خاطئة.

SB-11— تحيز المنتج والخطة والواجهة

فقط مستخدمو خطة أو جهاز معين يحددون اللوحة.

SB-12— تحيز الألفة بالعلامة التجارية

العملاء أو المعجبون أو النقاد أو الموظفون ممثلون بشكل زائد.

SB-13— الاحتيال والتحيز المتكرر

الشخص نفسه أو المشاركة التلقائية تكتسب تأثيرًا كبيرًا على النتيجة.

SB-14— تحيز استبعاد الصلاحية

غالبًا ما تُعتبر استجابات الذكاء الاصطناعي السلبية غير صالحة.

SB-15— تحيز الوزن

تغيير نموذج الوزن للمتغيرات المختارة بعد النتيجة أو التغيير المفرط في درجة النموذج في الاتجاه المرغوب.

SB-16— تحيز العينة الناجية

يتم عرض المستخدمين الذين يمكنهم إكمال المهمة بنجاح فقط كما لو كانوا يمثلون سكان المنتج الفعلي.

35. غرض الوزن

يمكن أن يخدم الوزن الأغراض التالية:

- تصحيح احتمالات الاختيار غير المتساوية
- إرجاع الإفراط في العينة إلى نسبة السكان الفعلية
- تقليل الفروق في معدلات الاستجابة
- معايرة العينة حسب المجموعات الهدف المعروفة
- دمج مصادر لوحات متعددة

لا يمكن للأوزان أن تخدم الأغراض التالية:

- إنشاء شخص غير موجود في إطار العينة
- تصحيح المتغيرات غير المقاسة بدقة
- جعل ملاحظة خاطئة أو زائفة حقيقية
- إنتاج النتيجة المطلوبة بعد الحصول على النتيجة
- تحويل خلية صغيرة واحدة إلى نتيجة موثوقة للبلد
- تحويل لوحة غير احتمالية إلى عينة احتمالية تلقائيًا

36. البنية الثقيلة

يمكن عرض الوزن النهائي كما يلي:

```
w_i= d_i× a_i^{NR}× g_i^{CAL}× p_i^{SRC}
```

هنا:

- di: الوزن الأساسي للتصميم أو وزن البداية
- aiNR: إعداد الاستجابة
- giCAL: معايرة السكان
- piSRC: مصدر اللوحة أو تعديل العينة غير المحتملة

في اللوحة الاحتمالية:

```
d_i= 1/π_i
```

هذا ممكن. πi الحقيقي غير معروف في اللوحة غير الاحتمالية. في هذه الحالة، الوزن المستخدم ليس وزن التصميم، بل الوزن المستند إلى النموذج أو الوزن المعاير. يجب الحفاظ على هذا التمييز في التقرير.

37. لا يمكن أن تعتمد الأوزان على النتيجة

لا يمكن استخدام المتغيرات التالية لإنشاء الأوزان:

- ما إذا كان رد الذكاء الاصطناعي صحيحًا أم لا
- ما إذا تم ذكر العلامة التجارية أم لا
- ما إذا كان هناك استشهاد
- ما إذا تم تقديم توصية
- ما إذا حدث خطأ جسيم
- ما إذا كانت المنظمة المدققة تحب النتيجة

متغيرات الوزن:

- قبل رؤية النتيجة،
- لأنه مرتبط بالسكان المستهدفين وآلية الاختيار

يجب اختياره. إذا كان نتيجة الاستجابة تحدد الوزن، فإن النتيجة تشكل نفسها.

38. متغيرات المعايرة

مناطق المعايرة المحتملة:

- البلد

- المنطقة
- اللغة
- العمر
- الجنس، إذا كان ضروريًا وأخلاقيًا
- التعليم
- الاستيطان الحضري/الريفي
- نوع الوصول إلى الإنترنت
- الجهاز
- تكرار استخدام الذكاء الاصطناعي
- الخطة المجانية/المدفوعة
- واجهة
- حالة الوصول

قد لا تكون الأهداف الموثوقة متاحة لكل متغير. العديد من متغيرات المعايرة:

- يمكن أن تخلق تباينًا مفرطًا في الوزن،
- حجم عينة فعّال منخفض،
- فرط التخصيص للنموذج

يجب تجميده قبل رؤية نتائج مجموعة المعايرة.

39. توازن الوزن

يجب مقارنة العينة قبل وبعد الوزن مع السكان المستهدفين. لكل حقل:

- الحصة المستهدفة
- حصة العينة الخام
- الحصة الموزونة
- الفارق المتبقي

يجب إظهاره. يتم تطبيق الوزن فقط على عدد النتائج ولا يمكن حفظ جدول التوازن.

40. تشخيصات الوزن

يجب أن يشمل التقرير الرئيسي الحقول التالية بقدر ما تكون ذات صلة:

- الحد الأدنى للوزن
- الوزن الأقصى
- الوزن الوسيط
- معامل تباين الوزن
- نسبة الوزن
- عدد الأوزان المقتطعة
- حجم العينة الخام
- حجم العينة الفعّال

- تأثير التصميم
- النتيجة غير المرجحة
- النتيجة المرجحة الأساسية
- نتائج الأوزان المعقولة البديلة

الرقم المرجح الفردي ليس سجل تحكم كامل بدون معرفات الوزن.

41. عرض الترجيح الاصطناعي

عرض المنهجية الاصطناعية / المشاركون، المعدلات، والنقاط أدناه افتراضية. التوزيع والتعداد المستهدف للسكان:

تكرار استخدام الذكاء الاصطناعي	حصة السكان المستهدفين	نسبة اللوحة الخام	الملاحظة الخام
استخدام منخفض أو غير متكرر	50%	10%	100
استخدام متوسط	30%	30%	300
استخدام كثيف	20%	60%	600
الإجمالي	100%	100%	1,000

معدلات النجاح الاصطناعية:

المجموعة	معدل النجاح	الاستجابة السابقة
استخدام منخفض أو غير متكرر	72%	72
استخدام متوسط	84%	252
استخدام كثيف	96%	576
الإجمالي	90%	900

النتيجة الخام:

```
900/1000=90%
```

41.1. معاملات المعايرة

استخدام منخفض:

```
0.50/0.10=5
```

استخدام متوسط:

```
0.30/0.30=1
```

استخدام مرتفع:

```
0.20/0.60 = 1/3
```

الاستجابة المرجحة:

```
72(5) + 252(1) + 576(1/3) = 804
```

إجمالي الوزن:

```
100(5) + 300(1) + 600(1/3) = 1,000
```

النتيجة المرجحة:

```
804/1000=80.4%
```

41.2. حجم العينة الفعال

```
n_eff= [100(5)+300(1)+600(1/3)]² / [100(5 ²)+300(1 ²)+600(1/3)²] ≈ 349
```

بطاقة النتيجة:

- الملاحظة الصالحة الخام: 1,000
- الدرجة الخام: 90 في المائة
- الدرجة الموزونة: 80.4 في المائة
- العينة الفعالة: حوالي 349
- السبب الرئيسي: التمثيل الزائد لمستخدمي الذكاء الاصطناعي بكثافة

التفسير الصحيح:

> تحتوي اللوحة على *1,000* من الملاحظات الصالحة؛ ومع ذلك، نتيجة لمعايرة السكان المستهدفين، انخفض التقدير إلى *80.4%*، وبسبب تباين الأوزان، أصبحت حجم العينة الفعالة حوالي *349.*

التعليق غير الصحيح: "تم إثبات الدقة العالمية لتكون 90% مع 1,000 من الأشخاص."

42. ما الذي لا يمكن للوزن تصحيحه؟

يجب ألا تكون المجموعات التالية موجودة في اللوحة التركيبية على الإطلاق:

- الأشخاص الذين يستخدمون برامج قراءة الشاشة
- المستخدمون في المناطق الريفية منخفضة النطاق الترددي
- الأشخاص الذين يستخدمون منتج الذكاء الاصطناعي لأول مرة
- مجتمع لغوي صغير

يغير الوزن فقط وزن المستخدمين في اللوحة. المجموعة غير المرصودة:

- ذكاء اصطناعي، إجابتك،
- كشف مشاكل الوصول,
- توزيع الخطأ

لا يمكن إنتاجه. لذلك:

> صفر التغطية يعني صفر المعلومات.

النتيجة الموزونة: قد تكون تنبؤًا أقرب إلى المجموعة المستهدفة، وليس ملاحظة كاملة للمجموعة المستهدفة.

43. تحليل الحساسية

يجب فحص اعتماد النتيجة الموزونة على نموذج وزن واحد. النتائج المرشحة:

- النتيجة الخام غير الموزونة
- النتيجة المركزة على التصميم
- النتيجة المعايرة ديموغرافيًا
- المعايرة مع إضافة استخدام الذكاء الاصطناعي
- النتيجة بعد تقليم الوزن

- النتيجة المعدلة بتأثير موفر اللوحة

إذا تغيرت النتائج بشكل جوهري:

تحذير حساسية الوزن

ينبغي إعطاؤه. لا يمكن تقديم نتيجة وزن واحدة "صحيحة" كحقيقة مطلقة.

44. تعديل الميل في اللوحة غير الاحتمالية

يمكن نمذجة ميل أعضاء اللوحة غير الاحتمالية لدخول اللوحة وفقًا للسكان المستهدفين. احتمال أن يكون شخص ما في اللوحة:

```
p̂_i= Pr(i ∈ panel | X_i)
```

يمكن تقديره. وزن الميل العكسي المرشح:

```
w_i^P= 1/p̂_i
```

يمكن استخدامه. هذه الطريقة:

- تستخدم فقط المتغيرات Xi المقاسة،
- تتأثر بالنموذج المحدد بشكل خاطئ,
- الاختيار غير المرصود لا يحل المشكلة،

يتطلب بيانات مرجعية قوية للهدف. لذلك، النتيجة:

يجب وسمها كالتالي:

تقدير لوحة غير احتمالية معايرة بالنموذج

45. تقديرات المناطق الصغيرة والنموذجية

قد تكون العينات المباشرة غير كافية في بعض خلايا الدول أو اللغات. يمكن لطرق القائمة على النموذج أن تستعير المعلومات من المجموعات المجاورة أو المماثلة. هذا يعني أن النتيجة:

- درجة الدولة المرصودة مباشرة،
- معدل المستخدم الخام

ليست كذلك. الحالة الصحيحة:

التقدير بمساعدة النموذج

تقدير المناطق الصغيرة

لم يتم ملاحظته مباشرة

يجب. التنبؤ المستند إلى النموذج لا يحول دولة دون ملاحظة مراقب الدولة إلى دولة ملاحَظة.

46. خصوصية المشاركين وإمكانية التحقق

GEO-1000 يتطلب إثبات لكل ملاحظة. هذا لا يعني أنه سيتم الكشف عن هوية المشارك علنًا. قد تتضمن لقطة الشاشة المعلومات التالية:

- الاسم

- البريد الإلكتروني
- صورة الملف الشخصي
- الدردشة السابقة
- خطة الحساب
- اسم المؤسسة
- معلومات شخصية حساسة

النظام التقاط NOMOS أو ما يعادله:

- يجب ألا يتم التقاط الحقول الشخصية غير الضرورية،
- يجب تنفيذ إخفاء تلقائي أو محكوم،
- يجب حفظ الأدلة الخام في مستودع آمن،

ينبغي تقديم السجلات المجهولة وقابلة للتحقق فقط للجمهور.

46.1. قائمة الأدلة العامة

السجل العام:

- معرّف المراقب،
- خلية البلد/اللغة،
- منتج الذكاء الاصطناعي،
- الوقت،
- تجزئة الأدلة،
- حالة التحقق

قد يتم حملها. لقطة شاشة كاملة:

- تم تحريرها،
- وصول مقيد،
- مفتوح للمراجع المستقلة

قد يكون. وجود ألف قطعة من الأدلة لا يتطلب الكشف العام عن هويات ألف شخص.

46.2. مخاطر الدول الصغيرة الحساسة

إذا كان هناك مشارك واحد فقط في الدولة الصغيرة:

- الجهاز،
- الوقت،
- اللغة،
- خطة

عند جمعهم معًا، قد يصبح من الأسهل إعادة التعرف على الشخص. يجب تقليل مستوى التفاصيل في التقارير العامة. يمكن الحفاظ على سلامة الأدلة في نظام يتم التحكم في الوصول إليه.

47. موافقة المشارك

يجب أن يعرف المشارك على الأقل ما يلي:

- أي مهمة سيؤديها

- أي البيانات ولقطات الشاشة التي سيتم جمعها
- ما هي المعلومات الشخصية التي لن يتم جمعها
- مدة تخزين البيانات
- من سيكون لديه الوصول
- كيفية نشر النتيجة
- كيفية تقديم المكافأة
- أن المشاركة طوعية
- الحق في الانسحاب أو الاعتراض
- لن يُطلب كلمة مرور الحساب
- لا يُتوقع انتهاك شروط منتج الذكاء الاصطناعي

يجب ألا يكون الموافقة مخفية في نص طويل وغير مفهوم. [K06; K22]

48. سلامة المشترك

من المشترك:

- تقديم معلومات صحية أو قانونية أو مالية شخصية للذكاء الاصطناعي،
- استخدام بيانات عملاء حقيقية،
- مشاركة معلومات سرية لصاحب العمل،
- إضعاف أمان الحساب،
- تجاوز حدود الاستخدام،
- انتهاك القانون أو شروط المنتج

لا يمكن طلب ذلك. يجب أن تحتوي المُدخلات التجريبية على سياق عام وغير شخصي قدر الإمكان.

49. يحمِل المشارك التنوع البشري، ليس الجودة

المستخدم غير المتمرس:

- يمكن أن يُكمل المهمة ببطء أكثر،
- أصعب،
- مع مزيد من الدعم

هذا الشخص ليس تلقائيًا مشاركًا منخفض الجودة. هم جزء حقيقي من الفئة المستهدفة. إذا كان تصميم البحث معقدًا جدًا بحيث لا يمكن إلا للأشخاص ذوي المهارات الرقمية العالية إكماله: قد لا تكون المشكلة في المشارك، بل في أداة القياس. البروتوكول:

- يجب أن يكون مفهومًا،
- متاحًا،
- طبيعيًا لغويًا،
- العبء التقني محدود

يجب أن تكون.

50. بطاقة النتيجة العامة لاختيار المشاركين

يجب أن يتضمن التقرير العام للمرشح المجالات التالية:

موارد اللوحة

احتمالية مشاركة اللوحة مشاركة المتطوعين المحترفين مشاركة الدعوة المفتوحة مشاركة الجامعة/المؤسسة مشاركة العملاء أو اللوحة التابعة

قمع المشاركة

الشخص في الإطار مختار مدعو مقبول بدأ مكتمل إنتاج ملاحظة صالحة

الأهلية والنزاهة

مستويات التحقق من الدولة مستويات التحقق من اللغة معدل تكرار الشخص معدلات التسجيل المشبوهة معدل الاحتيال المؤكد معدل الملاحظات غير الصالحة

تركيب اللوحة

تكرار استخدام الذكاء الاصطناعي الخطة والواجهة معرفة العلامة التجارية علاقة العميل/الموظف المستخدمون الجدد والعائدون للوحة تغطية إمكانية الوصول

تشخيصات الوزن

يمكن نشر ما يلي دون الكشف عن هوية المشاركين: الدرجات الخام والمرجحة؛ الحد الأدنى والأقصى للأوزان؛ معامل اختلال الأوزان؛ حجم العينة الفعّال؛ الحساسية تجاه الأوزان؛ وتحذيرات التغطية.

51. الأحكام الإلزامية المعيارية

CH08-N01

يجب أن تحمل كل ملاحظة صالحة GEO-1000 سلسلة اختيار المشاركين قابلة للتتبع من المجتمع المستهدف إلى سجل التحليل.

CH08-N02

لا يمكن تقديم الاختيار العشوائي داخل اللوحة على أنه اختيار عشوائي من المجتمع العالمي المستهدف.

CH08-N03

يجب شرح مصدر المشاركين، وطريقة الاختيار، وحالة اللوحة في سجل الطريقة العام.

CH08-N04

لا يمكن تعريفها كعينة احتمالية غير محتملة أو تطوعية من السكان ضمن اللوحة.

CH08-N05

لا يمكن إدراج المشاركين الذين اقترحتهم المؤسسة التي تم تدقيقها في اللوحة السكانية المستقلة الرئيسية دون توضيح.

CH08-N06

يجب تسجيل علاقات العملاء والموظفين والشركاء والمنافسين ومزودي الذكاء الاصطناعي بشكل منفصل.

CH08-N07

يجب التحقق من معايير أهلية المشاركين قبل رؤية استجابة الذكاء الاصطناعي.

CH08-N08

يجب تحديد أهلية البلد واللغة والمنتج والخطة والواجهة قبل تعيين المهمة المتعلقة بها.

CH08-N09

لا يمكن تكرار نفس الشخص كوحدة سكانية بسبب تعدد الحسابات أو الأجهزة أو اللغات أو اللوحات أو منتجات الذكاء الاصطناعي.

CH08-N10

يجب إجراء التحقق من الشخص الفريد وفقًا لمبادئ تقليل البيانات والخصوصية.

CH08-N11

لا يمكن اعتبار إشارة تقنية واحدة بمفردها كافية لاتخاذ قرار محدد تلقائيًا بأن المشارك مزيف.

CH08-N12

لا يمكن ربط مكافأة المشارِك بكون رد الذكاء الاصطناعي إيجابياً أو طويلًا أو يحتوي على مراجع أو مفيدًا للعلامة التجارية.

CH08-N13

يمكن ربط علاوة الجودة فقط بنزاهة البروتوكول المحددة مسبقًا.

CH08-N14

لا يمكن لنص الدعوة والمعلومات توجيه المشارك للبحث عن نتيجة إيجابية أو سلبية.

CH08-N15

يجب تعيين منتج الذكاء الاصطناعي أو عشوائيته قبل رؤية النتيجة الخاصة بالمُدخل وتسلسل المهمة.

CH08-N16

لا يمكن استخدام اختيار المنتج وصياغة المُدخل بناءً على تفضيل المشاركين دون توضيح في اللوحة السكانية الرئيسية.

CH08-N17

يجب تسجيل تهيئة اللوحة، التعرض للاختبار السابق، والمشاركة المتكررة.

CH08-N18

يجب تحليل نتائج المشاركين الجدد والمشروطين بشكل منفصل عند الضرورة.

CH08-N19

لا يمكن أن يستند تضمين المشارك في العينة أو احتفاظه بها أو استبعاده أو استبداله ببديل إلى محتوى استجابة الذكاء الاصطناعي.

CH08-N20

يجب الحفاظ على مخرجات الذكاء الاصطناعي غير الصحيحة أو السلبية أو المرفوضة أو النقدية وفقًا لبروتوكول صالح.

CH08-N21

يجب تقييم صلاحية المشاركين بشكل أعمى قدر الإمكان وقبل تقييم الدقة الجوهرية للإجابة.

CH08-N22

يجب الإبلاغ عن معدلات الدعوة والقبول والبدء والانتهاء والملاحظة الصحيحة بشكل منفصل.

CH08-N23

يجب أن يكون إعداد الاستجابة قائمًا فقط على المتغيرات المحددة والمهمة.

CH08-N24

لا يمكن للتوزين أن يجعل المستخدمين غير الموجودين في إطار العينة ممثلين.

CH08-N25

لا يمكن أن تعتمد متغيرات الوزن على نتيجة استجابة الذكاء الاصطناعي أو الدرجة أو المصلحة التجارية للمؤسسة المنظمة.

CH08-N26

لا يمكن تقديم أوزان لوحة غير احتمالية على أنها أوزان تصميم.

CH08-N27

يجب الإفراج عن متغيرات المعايرة والإجماليات المستهدفة قبل عرض النتائج.

CH08-N28

يجب تخزين النتائج غير الوزنية والنتائج الوزنية الرئيسية معًا.

CH08-N29

يجب أن تكون تشخيصات التوزين وحجم العينة الفعال مرئية في التقرير الرئيسي.

CH08-N30

إذا تم استخدام تقليم الأوزان، يجب الاحتفاظ بالنتيجة غير المقلصة وتحليل الحساسية.

CH08-N31

يجب الاحتفاظ بتأثيرات مزود اللوحة ومصدر المشارك كمتغيرات منفصلة.

CH08-N32

في لوحة متعددة المصادر، يجب إجراء فحوصات تكرار الشخص بين المصادر.

CH08-N33

يجب أن تشمل كشف الاحتيال إشارات متعددة، ومراجعة بشرية متناسبة، وعملية استئناف.

CH08-N34

لا ينبغي اعتبار استخدام VPN أو جهاز مشترك أو شبكة مؤسسية أو تقنية مساعدة احتيالاً تلقائيًا.

CH08-N35

لا يمكن تقديم تقدير المنطقة الصغيرة القائم على النموذج مباشرة مثل درجة دولة أو لغة مُلاحظة.

CH08-N36

يجب أن تحمل لقطات شاشة المشارك والأدلة تقليل البيانات الشخصية، والتحرير، والتحكم في الوصول.

CH08-N37

لا يمكن طلب كلمات مرور الحسابات أو سجل الدردشة الخاص أو المعلومات الحساسة غير الضرورية من المشاركين.

CH08-N38

لا يمكن طلب من المشاركين تجاوز قيود المنتج أو العمر أو الهوية أو البلد أو الحساب.

CH08-N39

لا يمكن تسجيل الأطفال أو المستخدمين الذين يحتاجون إلى حماية خاصة في اللوحة بدون بروتوكولات أخلاقية وأمنية منفصلة.

CH08-N40

يجب أن يكون لسجلات اختيار المشاركين وتوزانهم مالك بشري أو مؤسسة يمكن محاسبته.

52. أشكال الفشل

CH08-F01— عد العشوائية داخل اللوحة كعشوائية من العالم

يتم تقديم الاختيار العشوائي داخل اللوحة التطوعية كما لو كان اختيارًا عشوائيًا من سكان العالم.

CH08-F02— الامتثال لحصص العد كممثل

يُدعى أنه تم حل جميع الانحيازات السلوكية والتكنولوجية لأن حصص العمر والبلد قد تم تلبيتها.

CH08-F03— عد العملاء على أنهم عامة السكان

تُستخدم قائمة مستخدمي المؤسسة المراقبة كما لو كانت لوحة مستخدمين عالمية.

CH08-F04— إخفاء الموظفين والشركاء

يتم عرض المستخدمين المرتبطين بالمؤسسة كمشاركين عامين مستقلين.

CH08-F05— عد مزود لوحة واحد على أنه عالمي

لا يتم فحص تأثير المزود وسلوكية اللوحة المسجلة.

CH08-F06— احتساب الدعوة المفتوحة كعينة محتملة

تم الحفاظ على الاختيار الذاتي.

CH08-F07— تحديد البلد فقط من خلال IP

يتم تجاهل اختلافات السفر، وVPN، والشبكة المؤسسية، ومنطقة الحساب.

CH08-F08— احتساب الجنسية كبلد الاستخدام

يتم تصنيف الحالة الجغرافية الفعلية لتجربة المنتج بشكل خاطئ.

CH08-F09— احتساب إعلان اللغة كنهائي دون اختبار

يظل المشارك في خلية اللغة رغم أنه لا يفهم المُدخل أو الاستجابة.

CH08-F10— تحويل اختبار اللغة إلى اختيار نخبة

يبقى فقط الأشخاص المتعلمون تعليماً عالياً في العينة.

CH08-F11— احتساب عدد الحسابات كأشخاص

يزيد الشخص الواحد الوزن السكاني باستخدام عدة حسابات.

CH08-F12— اعتبار الحساب المشترك كشخص واحد

يتم تفسير الحساب الذي يستخدمه عدة أشخاص كمستخدم واحد.

CH08-F13— الدفع اعتماداً على النتيجة

يتم منح مكافأة على الإجابة الإيجابية أو الصحيحة أو المرجعية.

CH08-F14— كسر البروتوكول مع مكافأة السرعة

يأخذ المشارك لقطة شاشة أو يسرع في أداء المهمة قبل إكمال الرد.

CH08-F15— شرح النتيجة في الدعوة

يُطلب من المشارك العثور على خطأ أو نجاح.

CH08-F16— السماح للمشارك باختيار المنتج

المستخدمون الذين يختارون فقط منتج الذكاء الاصطناعي الذي يمكنهم الوصول إليه أو الذي يعجبهم يحددون اللوحة.

CH08-F17— إخفاء تأثير ترتيب صياغات المُدخل

الإجابات السابقة تؤثر على المهام اللاحقة.

CH08-F18— اعتبار تهيئة اللوحة استخدامًا طبيعيًا

يتم تقديم المستخدم الذي يرى الاختبار نفسه عدة مرات كمستخدم جديد.

CH08-F19— تجاهل مشاركة النتائج

نقل المشاركون المهام والاستجابات لبعضهم البعض.

CH08-F20— احتساب عدم الاستجابات كبيانات مفقودة

لا يتم فحص الفرق المنهجي للمشاركين غير الموجودين.

CH08-F21— احتساب فقدان اللوحة كتحسن في النتائج

تزداد الدرجات لأن المستخدمين الضعفاء أو عديمي الخبرة ينسحبون في المراحل التالية.

CH08-F22— إبطال الاستجابات السلبية

يتم استبدال المشارك لأن الذكاء الاصطناعي قدم استجابة غير صحيحة.

CH08-F23— التحقق من الاستجابة الصحيحة

حتى إذا قام المشارك بتحديث الاستجابة، يظل السجل محفوظًا لأن النتيجة صحيحة.

CH08-F24— تجاهل الروبوتات والمشاركة المتكررة

نفس الشخص أو الأتمتة ينتج عددًا كبيرًا من الملاحظات.

CH08-F25— الاستبعاد الجماعي بإشارة واحدة

تسبب الشبكة الخاصة الافتراضية أو الجهاز أو السرعة في اعتبار المستخدمين الحقيقيين مزيفين.

CH08-F26— إجراء مراجعة الاحتيال بناءً على النتائج

يتم فحص السجلات ذات الدرجات المنخفضة بشكل أكثر تكرارًا.

CH08-F27— إنشاء مجموعة غير مرئية بشكل كبير

يُزعم أن المستخدمين الذين ليسوا موجودين في العينة يتم تمثيلهم.

CH08-F28— اختيار المعايرة بعد النتيجة

يتم بعد ذلك اختيار متغيرات الوزن التي تنتج أفضل نتيجة مناسبة.

CH08-F29— حساب حصة اللوحة الخام كحصة سكانية

يحدد مستخدمو الذكاء الاصطناعي المكثفون النتيجة العالمية بالأرقام الخام.

CH08-F30— حساب الوزن غير العادي كوزن تصميم

يتم إجراء استدلال دقيق للسكان على الرغم من أن احتمال الاختيار الحقيقي غير معروف.

CH08-F31— إخفاء النتيجة غير المرجحة

تأثير نموذج الوزن غير مرئي.

CH08-F32— إخفاء العينة الفعالة

على الرغم من أن سجلات 1,000 الخام تحمل وحدات المعلومات 400–300، فإنه يتم المطالبة بالدقة الكاملة.

CH08-F33— إزالة المجموعات الضعيفة عن طريق تقليم الوزن

يتم تقليم المجموعات الصغيرة ذات الوزن العالي لأنها تقلل النتيجة.

CH08-F34— عد الدرجات المقدرة بناءً على النموذج

يتم إعطاء دولة دون مثال درجة محددة.

CH08-F35— الكشف عن الهوية من أجل الأدلة

يتم نشر لقطات الشاشة علنًا مع المعلومات الشخصية.

CH08-F36— جعل مشارك من دولة صغيرة قابل لإعادة التعريف

معلومات البلد والوقت والجهاز والخطة تكشف عن شخص واحد.

CH08-F37— يطلب من المشارك كلمة المرور أو الدردشة الخاصة

جمع الأدلة ينتهك أمان الحساب.

CH08-F38— احتساب مهارة رقمية منخفضة على أنها عدم لياقة

يتم إزالة المستخدمين المستهدفين الفعليين من العينة.

CH08-F39— الإخلال بالاختيار الطوعي بالمكافأة

المشارك تحت ضغط لا يمكنهم رفضه على نحو معقول.

CH08-F40— ترك اختيار المشارك للعميل

الكيان الذي يتم تدقيقه يحدد الأشخاص الذين يمكنهم تحديد نتيجة العينة.

53. إجراء التدقيق

الخطوة —1 رسم خريطة مصادر المشاركين

يتم تسجيل كل لوحة بشكل منفصل للشريك، الدعوة المفتوحة، ومصدر العميل.

الخطوة —2 تحديد حالة الاختيار

يتم تحديد الحالة المناسبة بين PSI-0 و PSI-5 لكل مصدر.

الخطوة —3 مقارنة التغطية مع الفئة المستهدفة

تحديد المستخدمين المستهدفين الذين لا يتواجدون في اللوحة.

الخطوة —4 حفظ احتمالية اختيار المشاركين أو النموذج

يتم شرح الاحتمال المعروف، الحصة، أو الاختيار الطوعي.

الخطوة —5 تثبيت قواعد المطابقة

العمر الدولة اللغة المنتج خطة الواجهة التفرد الاستخراج العلاقة الشروط تم إصدار إصدارات لها.

الخطوة —6 مراجعة نص الدعوة

يتم تقييم ما إذا كان هناك توجيه إيجابي أو سلبي.

الخطوة —7 فحص هيكل المكافأة

يتم التأكد من أن المكافأة مستقلة تمامًا عن النتيجة.

الخطوة —8 التحقق من دولة المشارك، واللغة، والوصول إلى المنتج

يتم تعيين مستويات تحقق تحمي الخصوصية.

الخطوة —9 إجراء فحص الإنسان الفريد

يتم فحص التكرارات داخل المصدر وبين المصادر.

الخطوة —10 تسجيل مجالات العلاقة والألفة

يتم تحديد معرفة العميل والموظف والمنافس ومزود الذكاء الاصطناعي والعلامة التجارية.

الخطوة —11 توزيع المهام وفقًا لخطة التعيين

يتم التحقق من المنتجات وصياغات المُدخل والطابور والتخصيصات الموجبة.

الخطوة —12 تسجيل تهيئة اللوحة

يتم فصل المستخدمين الجدد والمتكررين والمشاركين مسبقًا في نفس الاختبار.

الخطوة —13 إعداد قمع المشاركة

يتم حساب أرقام الدعوة والقبول والبدء والإكمال والصلاحية.

الخطوة —14 إخفاء مراجعة الصلاحية

يتم تحديد صحة البروتوكول دون أن يُنظر إلى درجة استجابة الذكاء الاصطناعي قدر الإمكان.

الخطوة —15 فحص إشارات الاحتيال

يتم استخدام إشارات متعددة وتقييم بشري.

الخطوة —16 تحليل عدم استجابتي وخسارة اللوحة

يتم مقارنة الخصائص المعروفة لأولئك الذين يشاركون وأولئك الذين لا يشاركون.

الخطوة —17 توليد أوزان التصميم والمواءمة

يتم التعامل مع المصادر الاحتمالية وغير الاحتمالية بطرق منفصلة.

الخطوة —18 حساب تشخيصات الوزن

توزيع الوزن تأثير التصميم العينة الفعالة يتم حساب تأثير التشذيب.

الخطوة —19 إجراء تحليل الحساسية

يتم مقارنة النتائج الخام والرئيسية الثقيلة والبديلة المعقولة.

الخطوة —20 إنشاء سجل الخصوصية والأدلة

الدليل الكامل موجود في مستودع آمن؛ تُحفظ هويات الأدلة المجهولة في سجل عام.

الخطوة —21 نشر بطاقة الانتخابات العامة

مصادر اللوحة، قمع المشاركة، الانحيازات، وتأثير الوزن يتم جعلها مرئية.

54. الأدلة المطلوبة

سجل الفئة المستهدفة؛ إطار العينة؛ مقدمو اللوحات؛ مصادر المشاركين؛ حالة الاختيار؛ فئة اللوحة الاحتمالية أو غير الاحتمالية؛ طريقة تسجيل اللوحة؛ نص الدعوة؛ قواعد قبول المشاركين؛ التحقق من بلد المشارك؛ سجل كفاءة اللغة؛ سجل الوصول إلى منتج الذكاء الاصطناعي؛ الخطة و واجهة المستخدم؛ التحقق من الشخص الفريد؛

فحوصات التكرار عبر المصادر؛ معرفة العلامة التجارية؛ تكرار استخدام الذكاء الاصطناعي؛ العلاقة مع الجهة التي تخضع للتدقيق؛ العلاقة مع مزود الذكاء الاصطناعي؛ تعويض المشاركين؛ شروط المكافأة؛ طريقة تعيين المهام؛ سجل العشوائية والتقسيم؛ ترتيب المنتجات والتعليمات؛ تكييف اللوحة؛ المشاركة التجريبية السابقة؛ عدد المدعوين؛ عدد الذين وافقوا؛ الرقم الذي يبدأ المهمة؛ الرقم الذي يكملها؛ عدد الملاحظات الصالحة؛ أسباب عدم الاستجابة؛ تآكل اللوحة؛ أسباب عدم الصلاحية؛ ومؤشرات الاحتيال.

سجلات المراجعة والاعتراض الأوزان التصميمية إعدادات عدم الاستجابة متغيرات المعايرة نموذج الميل، إذا وُجد سجل تقليم الوزن حجم العينة الفعّال نتائج الحساسية تأثير مزود اللوحة الدرجات غير الموزونة والموزونة سجل الموافقة سياسة تخزين البيانات والوصول إليها قواعد التنقيح بيان الأدلة العامة الشخص أو المؤسسة المسؤولة

55. قائمة تدقيق التدقيق

من أي مصادر جاء المشاركون؟ هل كان كل مصدر عشوائيًا، طوعيًا، أم دعوة مفتوحة؟ هل تم تقديم العشوائية داخل اللوحة مثل عشوائية سكان العالم؟ هل تدخلت المؤسسة المشرفة في اختيار المشاركين؟ هل المشاركون من العملاء والموظفين منفصلون؟ هل تم توثيق العلاقات مع مزود الذكاء الاصطناعي أو المنافسين؟ كيف تم التحقق من أهلية البلد؟ هل إجادة اللغة على مستوى الاستخدام الفعلي؟ هل تم استخدام المنتج الصحيح للذكاء الاصطناعي والخطة والواجهة المناسبة؟ كم عدد الحسابات أو اللوحات التي شارك فيها نفس الشخص؟ هل تم إجراء التحقق المتبادل بين المصادر؟ هل نص الدعوة وجه النتيجة؟ هل تعتمد المكافأة على محتوى استجابة الذكاء الاصطناعي؟ هل قللت حوافز السرعة من جودة البروتوكول؟ هل كانت مهمة التعيين عشوائية أم بناءً على القواعد المسبقة؟ هل اختار المشارك المنتج أم قام بإدخال المُدخل بنفسه؟

هل كان المنتج والطلب الفوري متوازنين؟ هل تم تسجيل تكييف اللوحة؟ هل شارك المشاركون نتائج المهام؟ هل يمكن رؤية مسار الدعوة–القبول–البدء–الاكتمال–الصلاحية؟ هل يختلف غير المستجيبين في الخصائص المعروفة؟ هل يؤثر انسحاب المشاركين من اللوحة على اتجاه الوقت؟ هل تم اتخاذ قرار الصلاحية قبل تقييم استجابة الذكاء الاصطناعي؟ هل تم اعتبار الردود السلبية أو النقدية غير صالحة في كثير من الأحيان؟ هل يحتوي اكتشاف الاحتيال على إشارات متعددة؟ هل تسبب استخدام VPN أو أجهزة مشتركة في الاستبعاد التلقائي؟ هل تم توفير مسار للاستئناف للمشارك؟ هل تم تعريف متغيرات الوزن قبل النتيجة؟ هل تم تصنيف أوزان اللوحة غير المحتملة بشكل صحيح؟ هل المجموعات المستهدفة موثوقة؟ هل تُعرض النتائج غير الموزونة والموزونة معًا؟ هل حجم العينة الفعّال واضح؟ هل تم عرض تأثير تقليم الأوزان؟

هل تم إظهار مجموعات لم يتم ملاحظتها على أنها ممثلة بشكل كبير؟ هل تم تقديم التنبؤات المستندة إلى النماذج كما لو كانت ملاحظات مباشرة؟ هل هناك تفاصيل شخصية في لقطات الشاشة الخاصة بالمشاركين؟ هل يمكن إعادة تحديد هوية المستخدمين من دول صغيرة؟ هل تم طلب كلمات مرور الحسابات أو الدردشات الخاصة؟ هل موافقة المشاركين واضحة ومفهومة؟ هل هناك مالك مسؤول بوضوح عن اختيار المشاركين وسجلات التوزين؟

56. الاعتراضات والردود

الاعتراض 1— "إذا استخدمنا عينة بحثية محترفة، لماذا نزال نتحدث عن التحيز؟"

يمكن للوحة مهنية أن تعزز جودة البيانات وإدارة الميدان. كما أن أعضاء اللوحة هم:

- تطوعي،
- معتاد على البحث،
- مختلف في استخدام التكنولوجيا

من الممكن. الاحترافية لا تقضي على الانحياز. إنها تجعل القياس والإدارة أسهل.

اعتراض 2— "إذا تم تلبية حصص العمر والبلد، لماذا لا يُعتبر الفريق تمثيليًا؟"

المشاركون:

- تكرار استخدام الذكاء الاصطناعي،
- الجهاز،
- الخطة،
- الوعي بالعلامة التجارية،
- المهارة الرقمية

قد يختلف فيما يتعلق بالصيانة. الحصص توازن فقط المناطق التي تغطيها.

اعتراض 3— "ألا يمكن للأوزان تصحيح هذه الفروقات؟"

يمكنه تصحيح الفروقات جزئيًا عندما تكون هناك إجماليات هدف مقاسة وموثوقة. لا يمكنه تصحيح الفروقات بشكل قطعي التي لم يتم ملاحظتها أو قياسها أبدًا.

الاعتراض 4— "عملاء المؤسسة هم مستخدمون حقيقيون؛ لماذا لا يجب تضمينهم في اللوحة الرئيسية؟"

يمكن أخذهم. ومع ذلك، يجب أخذهم بالنسبة الفعلية وحالة العلاقة التي يحملونها في السكان العامين. قائمة العملاء الكاملة ليست لوحة للسكان العام.

الاعتراض 5— "لماذا سيكون هناك مشكلة إذا كان المشارك يعرف العلامة التجارية؟"

لا يجب أن تكون مشكلة. فهم جزء من مستخدمي النظام الطبيعيين. إذا كان الإلمام غير متوازن، فقد تختلف النتيجة. لذلك، يتم تسجيلها وتحليلها عند الضرورة.

اعتراض 6— "المستخدمون الكثيفون للذكاء الاصطناعي يؤدون المهمة بدقة أكبر؛ ألن يكون من الأفضل اختيارهم؟"

قد يكون اختبار البروتوكول أسهل. ومع ذلك، إذا أردنا قياس تجربة السكان الواقعية، فإن المستخدمين غير المتمرسين مهمون أيضًا. إذا كانت المهمة معقدة جدًا، فيجب تصحيح أداة القياس. لا ينبغي تبسيط السكان.

اعتراض 7— 'إذا أساء المشارك فهم الاستجابة، لماذا يجب علينا الاحتفاظ بالملاحظة؟'

مهمة المشارك ليست تقييم الرد، بل التقاطه بشكل صحيح. إذا تم تطبيق بروتوكول الالتقاط بشكل صحيح، يتم الحفاظ على ردود الذكاء الاصطناعي. تقييم المواد هو مسؤولية المحكمين.

اعتراض 8— "ألا ينبغي استخدام فلتر تلقائي صارم لمنع المشاركة المزيفة؟"

الفلترة التلقائية مفيدة. إذا أصبحت القرار النهائي الوحيد، فقد تستبعد المستخدمين الحقيقيين بشكل غير عادل. هناك حاجة إلى إشارات متعددة ومراجعة بشرية متناسبة.

اعتراض 9— 'إذا لم نجعل ألف لقطة شاشة عامة، هل يمكن اعتبار الطريقة مثبتة؟'

صور كاملة للشاشة:

- المعلومات الشخصية،
- أمان الحساب,
- إعادة التعريف

قد تحمل خطراً. للجمهور:

- أمثلة معدلة،
- ملاحظات العرض،

- تجزئات النزاهة،
- يمكن توفير الوصول إلى التدقيق المستقل

الإثباتية ليست كشف الهوية.

اعتراض **10—** "إذا كانت النتيجة الموزونة مختلفة جداً عن النتيجة الأولية، أيهما صحيح؟"

تجيب نتيجتان عن سؤالين مختلفين. النتيجة الأولية هي نتيجة الأشخاص في العينة. النتيجة الموزونة هي تقدير السكان الهدف. يشير الفرق الكبير إلى أن تكوين العينة مهم. يجب نشر كلتا النتيجتين.

اعتراض **11—** "إذا انخفضت العينة الفعالة إلى **349**، هل كان جمع **1,000** ملاحظات مضيعة للوقت؟"

قد تكون المجموعات المفرطة في أخذ العينات قد عززت تحليل الفئات الفرعية. ومع ذلك، قد تكون حساسية التنبؤ العالمي أقل من العدد الخام لـ .1,000 لا ينبغي إخفاء هذه الحقيقة.

اعتراض **12—** "هل يمكن عدم استخدام لوحة غير الاحتمالية على الإطلاق؟"

يمكن استخدامها. على وجه التحديد:

- تجريبية،
- لغة منخفضة الموارد،
- دولة صغيرة،
- اكتشاف سريع

إنها مفيدة لـ. يجب نشرها بالحالة الصحيحة والقيود المناسبة.

الاعتراض **13—** "أليس من الأكثر موثوقية اختيار جميع المشاركين من أشخاص جدد؟"

قد يكون قويًا لتقدير السكان العرضي. المستخدمون المتكررون ذو قيمة لقياس التغير الطولي. يجب الحفاظ على أدوار اللوحة الجديدة والمتكررة منفصلة.

اعتراض **14—** "ألا يزيد جمع البيانات من هذا العدد الكبير من المشاركين من خطر الخصوصية؟"

يمكنه ذلك. لذلك:

- تقليل البيانات
- هوية مجهولة
- تحرير,
- التحكم في الوصول
- مدة التخزين

هي أجزاء إلزامية من البروتوكول. لا ينبغي جمع المعلومات غير الضرورية.

البند الشائع في القسم **59**

أسهل رقم في دراسة GEO:

1,000

هذا ممكن. أصعب سؤال هو:

> *أي 1,000؟*

يمكن لمجموعتين لهما نفس الحصص حسب الدولة والعمر أن تنتجا نتائج مختلفة تمامًا. في إحداهما:

- المستخدمون المكثفون للذكاء الاصطناعي،
- المشتركون في الخطط المدفوعة،
- عمال التكنولوجيا

هم الأغلبية. في الأخرى:

- المستخدمون غير المتكررون،
- مستخدمو الهواتف المحمولة،
- الأشخاص ذوو المهارات الرقمية المنخفضة

إنه أكثر وضوحًا. كلاهما من أشخاص 1,000. إنها ليست نفس العينة السكانية. إذا أخفيت من أين جاء المشاركون ونشرت فقط النتيجة المرجحة، فإنك تحول الترجيح إلى ستار من الثقة وليس إلى طريقة للتحكم. لا يمكنك ترجيح شخص غير موجود في العينة. لا يمكنك تغيير إجابة المستخدم لأنها خاطئة. الإجابة الخاطئة ليست فشلًا للبحث. إنها الحدث الذي يحاول البحث قياسه. إذا اتبع المشارك البروتوكول بشكل غير صحيح رغم إعطاءه الإجابة الصحيحة، لا يمكنك الحفاظ على السجل. الإجابة الصحيحة لا تجعل من الملاحظة المزيفة حقيقة. يمكنك دفع مبلغ أكبر لشخص واحد. لكن ليس لإنتاج الإجابة التي تريدها. يمكنك الدفع وفقًا لظروف المعيشة في الدولة. لكن لا يمكنك شراء استعداد المشارِك. يمكن أن تكون ألف لقطة شاشة دليلًا. لست مُلزَمًا بالكشف عن هوية ألف شخص للعامة.

يمكن حماية نزاهة الأدلة وخصوصية الإنسان معًا. لذلك، فإن قانون القياس الثامن لـ NOMOS هو كما يلي:

> تُحدّد قوة العينة ليس فقط بعدد الأشخاص الذين تتكوّن منهم، بل أيضًا بسبب وكيفية اختيار هؤلاء الأشخاص.

القانون التاسع هو كما يلي:

> إعادة الوزن تُعيد توازن السكان المرصودين؛ لكنها لا تخلق الإنسان غير المرصود.

القانون العاشر هو:

> يمكن أن يكون جواب الذكاء الاصطناعي غير الصحيح بيانات صالحة؛ ويمكن أن يكون جواب الذكاء الاصطناعي الصحيح بيانات غير صالحة.

القانون الحادي عشر هو كما يلي:

> يجب منح المكافأة للمشارك ليس لاكتشاف النتيجة المرغوبة، بل لإثبات النتيجة المرصودة دون تغييرها.

القانون الثاني عشر هو كما يلي:

> إمكانية الإثبات لا تتطلب من المشارك التنازل عن الخصوصية.

وصية الفصل 8 لـ NOMOS

> لا تحضر لي ألف شخص بمفردهم. / أرني من أي عالم يأتي هؤلاء الألف شخص.
>
> لا تعامل عينة المتطوعين كعينة عشوائية من السكان، أو قائمة العملاء كالجمهور العام، أو الموظف كمستخدم مستقل.
>
> لا تظن أن جميع الفروقات التي لم تقم بقياسها قد تحسنت لمجرد أن الحصص الخاصة بالبلد والعمر قد تم تحقيقها.
>
> لا تضاعف شخصًا واحدًا بعدد حساباته أو أجهزته أو لغاته أو عضوياته في اللوحات.
>
> لا تضع أولئك الذين يحبون العلامة التجارية، وأولئك الذين يكرهون العلامة التجارية، وأولئك الذين لا يعرفون العلامة التجارية على الإطلاق في نفس الصندوق غير المرئي.

ادفع للمشارك ليس ليجد الإجابة الصحيحة، بل ليحفظ أول إجابة مناسبة يراها دون تغييرها.

لا تغير المستخدم عندما يقدم الذكاء الاصطناعي إجابة خاطئة. / لا تتحقق مما إذا كانت الإجابة صحيحة عندما يخالف المستخدم البروتوكول.

لا تجعل شخصًا غير موجود في العينة يبدو وكأنه مهم أو موجود.

لا تخفِ النتيجة الخام. / لا تعلن فقط النتيجة المرجحة كحقيقة. / أظهر العينة الفعّالة. / أظهر حساسيتها.

لا تعلن عن شخص يستخدم شبكة افتراضية خاصة على أنه مزيف، أو شخص يستخدم جهازًا مشتركًا على أنه بوت، أو شخص يستخدم تقنية مساعدة على أنها أتمتة.

تحقيق في الاحتيال. / لكن لا تحول التحقيق نفسه إلى تحيز استبعادي جديد.

احتفظ بلقطة شاشة. / احمِ اسم الشخص، حسابه، وحياته الخاصة.

أولاً، حدد السكان. / ثم اعرض الإطار. / ثم اشرح من اخترت، ومن لم يحضر، ومن لم يتمكن من إكماله، ولماذا استبعدت شخصًا ما. / وفقط بعد ذلك، قرر مقدار ما يمكن لدرجتك أن تعبر عن الإنسانية.

جملة ختام الفصل

في *GEO-1000*، المشارك هو أكثر من مجرد شخص يحمل ردًا. المشارك هو وحدة أخذ عينات بشرية يجب تبرير قدرتها على تمثيل الآخرين في السكان المستهدفين من خلال مسار الاختيار، والوزن، والاستقلالية، وإجراءات حماية الخصوصية.

النواة المعيارية

يجب أن تحتفظ كل ملاحظة **GEO-1000** بمسار قابل للتدقيق لاختيار المشاركين من السكان المستهدفين وإطار أخذ العينات عبر الاختيار، والدعوة، والموافقة، والتعيين، والإكمال، والتحقق من البروتوكول، والاستبعاد، والتوزين النهائي. لا يجوز تمثيل الاختيار العشوائي داخل اللوحة كاختيار عشوائي من السكان المستهدفين العالميين ما لم يكن آلية شمول كامل السكان تدعم هذا الادعاء. يجب أن يظل تجنيد المشاركين، والتعيين، والتعويض، والتحقق، والاستبدال، والاستبعاد، والتوزين مستقلاً عن ما إذا كانت استجابة الذكاء الاصطناعي إيجابية، سلبية، صحيحة، غير صحيحة، مرفوضة أو حرجة، أو مؤاتية تجارياً. يجب أن يظل الأشخاص والحسابات والأجهزة والجلسات واللغات والعضويات في اللوحة والملاحظات المتعلقة بمنتجات الذكاء الاصطناعي وحدات مميزة. يجب تحديد مصادر المشاركين الاحتمالية، والموافقة الطوعية، والدعوة المفتوحة، والمقدمة من العميل، والمؤسسية، والهجينة بشكل منفصل. قد يقلل الوزن من الاختلال الملحوظ، لكنه لا يجب أن يُمثل على أنه يخلق ملاحظات للفئات الغائبة عن إطار العينة أو يصحح الاختيار غير المقاس بيقين. يجب أن تظل النتائج غير المرجحة، والنتائج المرجحة، وتشخيصات الوزن، والفجوات في التغطية، وعدم الاستجابة، وحجم العينة الفعّال، وتحليلات الحساسية مرئية. يجب أن تستخدم قرارات الاحتيال والتكرار مراجعة متعددة الإشارات متناسبة ويجب ألا تُعامل استخدام الشبكات الخاصة الافتراضية (VPN)، أو الأجهزة المشتركة، أو الشبكات المؤسسية، أو أدوات الوصول تلقائيًا على أنها تزييف. يجب أن تبقى الأدلة المقدمة من المشاركين قابلة للتحقق دون الكشف عن الهوية أو بيانات الحساب أو المحادثات الخاصة أو الموقع الدقيق بشكل غير ضروري. يجب أن يتم توثيق كل قرار بشأن اختيار المشاركين وتوزينهم وإسناده إلى إنسان أو منظمة مسؤولة.

الجزء الثالث

إجراء التجربة وسلسلة الأدلة

حد الفصل

الفصول الثمانية الأولى أرست ما يلي: الاستجابة الفردية ليست درجة GEO؛ التمثيل هو توزيع؛ يجب التعرف على الكيان الذي تم تدقيقه قبل ملاحظة النتائج؛ يجب أن يظل التمثيل الرسمي وواقع الكيان الموثق منفصلين؛ يجب تعريف مجموعة المستخدمين المؤهلة لكل منتج ذكاء اصطناعي؛ يجب تخصيص الملاحظات الصالحة 1,000 من خلال تصميم عينة قائم على السكان؛ يجب أن تظل البلدان واللغات الأصغر مرئية من خلال لوحات تكميلية؛ ويجب أن تظل عملية اختيار المشاركين والاستبعاد والوزن مستقلة عن محتوى الاستجابة. يجب علينا الآن تعريف اعيناتب الآخر من التجربة:

> أي نظام ذكاء اصطناعي يقيسه المشاركون بالفعل؟

اسم المزود؟ نموذج أساسي؟ منتج محادثة؟ واجهة ويب مجانية؟ تطبيق جوال مدفوع؟ جلسة مع تمكين البحث على الويب؟ وضع استجابة لا يستشهد بالمصادر؟ مساحة عمل للشركات مدعومة بالبيانات الداخلية؟ حساب مع تمكين الذاكرة؟ محادثة مؤقتة أم نظيفة؟ هل واجهات المستخدم المختلفة لنفس المزود هي نفس النظام؟ هل يجب قياس منتجين يستخدمان نفس عائلة النماذج بنفس الطريقة؟ حتى لو أظهر المنتج نفس اسم النموذج على الشاشة:

- الاسترجاع،
- اختيار المصادر،
- قواعد الأمان،
- التخصيص،
- الوصول إلى الأدوات،
- الميزات التجريبية

إذا تغير، هل يمكن مقارنة النتائج؟ عندما يتم تحديث منتج الذكاء الاصطناعي في منتصف موجة القياس، هل يمكن اعتبار الردود الأولى والأخيرة في الموجة أنها جاءت من نفس النظام؟ مستخدم لديه:

- البحث على الويب مفعل،
- الذاكرة مفعلة،
- لديه محادثات سابقة،
- خطة مدفوعة

الردود؛ مستخدم آخر:

- البحث متوقف،
- الذاكرة متوقفة،
- جلسة جديدة،
- خطة مجانية

هل يمكن دمجه في نفس الخلية مع إجابته؟ لا. مبدأ القياس الذي وضعته الأقسام السابقة واضح: لا يوجد "ذكاء اصطناعي" واحد؛ ففهرس البحث، ونظام الاسترجاع، وإعادة الترتيب، ونموذج اللغة، وطبقة الاقتباس، ونظام الأمان قد تتصرف بشكل مختلف. نفس الإطار يتطلب أن يحتوي كل نتيجة قابلة للمراجعة على:

- النموذج أو المنتج،
- تاريخ،
- البلد،
- اللغة،
- مجموعة الاستعلامات،
- عدد التكرارات،
- سجل القياس

هذا القسم يطبق تلك المبادئ داخل سجل نظام الذكاء الاصطناعي. هذا القسم:

- التمييز بين المزود، المنتج، النموذج، و واجهة المستخدم
- مثال منتج الذكاء الاصطناعي,
- مستويات ثقة هوية النموذج،
- استرجاع وتكوينات الأدوات،
- حالات الذاكرة والتخصيص،
- الفروق في الخطة والواجهة والميزات التجريبية،
- تغييرات المنتج،
- تكوين النظام في وقت القياس،
- مستويات القابلية للمقارنة بين المنتجات،
- مقارنات نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة،
- انحراف الإصدار مع مرور الوقت،

يحدد. هذا الفصل لم يتطرق بعد إلى:

- نظام الترجمة الكاملة ومعادلة الدلالات للمحفزات،
- جميع تفاصيل لوحات المستخدمين المسيطر عليها والطبيعية,
- برمجيات NOMOS وهندسة البرمجيات الخاصة بها،
- تقييم الادعاءات الذرية للردود،
- المعاملات النهائية للنموذج وصيغة تقييم NOMOS

لا تُنهى. السؤال الرئيسي في القسم 9 هو:

> قبل مقارنة ردين من نفس صياغة المُدخل، كيف نثبت أن هذه الردود تأتي فعليًا من ظروف منتج ذكاء اصطناعي قابلة للمقارنة؟

التحدي NOMOS

مستخدمان لنفس المزود يسألان نفس السؤال: "ما نوع الشركة Apple.com؟"، المستخدم الأول:

- يستخدم المنتج المجاني على الويب،
- يفتح دردشة جديدة،
- وقد أوقف الذاكرة.

- لا يمتلك الوصول إلى الويب،

يتلقى إجابة فقط باستخدام قدرة المعرفة المدمجة في المنتج. المستخدم الثاني:

- باستخدام منتج محمول مدفوع،
- الذاكرة مفعلة،
- قد تحدث سابقًا عن Apple،
- البحث على الويب مفعل،

يتلقى إجابات مع مصادر حالية واستشهادات. الإجابة الأولى: "Apple هي شركة تكنولوجيا تنتج الإلكترونيات الاستهلاكية." الإجابة الثانية: "Apple هي شركة تكنولوجيا عالمية تقدم الأجهزة والبرمجيات والخدمات الرقمية. يمكنك الاطلاع على معلومات الشركة الحالية في هذه المصادر." كلا الإجابتين تأتيان من نفس المزود.

هل هذه نتيجة لنفس النظام؟ لا. الآن فكر في مستخدمين آخرين. كلاهما يرى نفس اسم النموذج على الشاشة. في بلد أحد المستخدمين، الميزة على الويب مفعلة. في بلد المستخدم الآخر، الميزة مطفأة. في إجابة أحد المستخدمين، الروابط المصدرية مرئية. واجهة المستخدم الأخرى لا تظهر الاقتباسات حتى لو كانت تستخدم نفس المصادر. هناك نفس تسمية النموذج.

هل هناك نفس خلية القياس؟ لا. الآن ضع في اعتبارك منتجين من مزودين مختلفين. كلاهما:

- دردشة جديدة
- الذاكرة متوقفة،
- نفس اللغة،
- نفس الدولة،
- نفس نافذة الوقت،
- تم تمكين الوصول إلى الويب,
- نفس إصدار صياغة المُدخل

يتم قياسها تحت هذه الظروف. وهي مبنية على نماذج أساسية مختلفة. هل يمكن مقارنتها؟ نعم، يمكن مقارنتها من حيث نتيجة مستخدم محددة. لأن موضوع المقارنة هو:

> وليس ما إذا كانت تستخدم نفس النموذج الأساسي، بل ما التمثيل الذي تنتجه تحت نفس شرط المستخدم المحدد.

الآن، دع نسخة المنتج تبقى دون تغيير لأول 500 من ملاحظات موجة القياس. خلال الملاحظات 500 التالية، يقوم المزود:

- بتغيير نظام الاسترجاع،
- قواعد الأمان،
- واجهة المستخدم

يقول التقرير وحده: "اختبر مستخدمو 1,000 نفس منتج الذكاء الاصطناعي." في الواقع، قد يكون تم قياس حالتين مختلفتين للمنتج. النص الأول من هذا القسم هو كما يلي:

> اسم المزود غير كافٍ لتحديد نظام الذكاء الاصطناعي المقاس.

ينص حكمه الثاني على:

> اسم النموذج ليس هو كل سلوك المنتج الذي يصل إلى المستخدم.

تنص الفقرة الثالثة منه على:

> نفس المزود، نفس تسمية النموذج، أو نفس الشعار لا يضمنان القابلية للمقارنة.

تنص المادة الرابعة على:

قد تكون المزودات المختلفة قابلة للمقارنة من حيث نتائج المستخدم إذا كانت شروط القياس متطابقة بشكل كافٍ.

تنص أحكامه الخامسة على:

إذا تغير المنتج ماديًا أثناء القياس، فلا يمكن الإبلاغ عنه كنظام ثابت واحد.

1. غرض الفصل

الغرض من هذا القسم هو تحديد منتج الذكاء الاصطناعي الذي تم تدقيقه ضمن GEO-1000 مع جميع تكويناته المادية وتحديد الشروط التي يمكن بموجبها مقارنة نتائج المنتجات المختلفة. يُوحد هذا القسم التمييزات التالية:

- المزود ومنتج الذكاء الاصطناعي
- منتج الذكاء الاصطناعي والنموذج الأساسي
- عائلة النموذج والنشر المحدد
- تصنيف النموذج وهوية النموذج الموثق
- واجهة المستخدم وواجهة برمجة التطبيقات (API)
- المنتج الويب والمنتج المحمول
- الخطة المجانية والخطة المدفوعة
- المنتج الاستهلاكي والمنتج المؤسسي
- الإجابة الأساسية للنموذج والإجابة المحسّنة باسترجاع المعلومات
- البحث المباشر على الويب مع الاسترجاع
- عرض الاقتباس مع استخدام المصدر
- استخدام الأداة مع الوصول إلى الأداة
- سياق الجلسة مع الذاكرة
- سياسة النظام مع التعليمات المخصصة
- الواقع المادي مع التخصيص
- دردشة مؤقتة مع محادثة جديدة
- حالة المنتج مع مرور الوقت مع اسم المنتج الثابت
- تغير السلوك الملحوظ مع التحديث المعلن من مزود الخدمة
- مقارنة بين نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة
- مثال على منتج قابل للمقارنة ضمن نفس عائلة المنتجات
- نتيجة المستخدم مقابل القدرة المجردة داخل النموذج
- غياب المنتج مقابل فشل المنتج
- الخطأ الفني مقابل الرفض الفعلي
- نفس صياغة المُدخل مقابل نفس التجربة
- نفس المتوسط مقابل نفس سلوك النظام
- لقطة فورية للمنتج مقابل أدائه على المدى الطويل

في نهاية هذا القسم، يجب أن تكون كل مراجعة GEO-1000 قادرة على الإجابة على الأسئلة التالية:

أي مزود وأي منتج، وأي واجهة، وأي خطة، وأي تكوين تم قياسه؟

إلى أي مدى تم التحقق فعليًا من اسم الطراز الذي ظهر على الشاشة؟

ما حالة الويب، والاسترجاع، والاستشهاد، والذاكرة، والتخصيص، والأدوات؟

هل ظل المنتج كما هو طوال فترة القياس؟

على أي مستوى من المقارنة تم وضع نتيجتي المنتج جنبًا إلى جنب؟

هل الشيء الذي يتم مقارنته هو الطراز الأساسي، المنتج الاستهلاكي، واجهة المستخدم، أم سلسلة النظام بأكملها؟

2. الأحكام المعيارية المركزية

يجب ربط كل ملاحظة *GEO-1000* ليس فقط باسم المزود أو النموذج؛ بل يجب ربطها بسجل مسجل نظام الذكاء الاصطناعي المؤرّخ الذي يحدد مثيل منتج الذكاء الاصطناعي وقت القياس.

يجب أن يحمل مثيل منتج الذكاء الاصطناعي على الأقل الحقول التالية قدر الصلة:

- المزود
- اسم المنتج
- واجهة المستخدم
- نوع المنتج
- الخطة أو فئة الحساب
- العلامة النموذجية الظاهرة
- مستوى الثقة في معرف النموذج
- إصدار المنتج أو النشر
- الدولة والموقع
- اللغة
- الوصول عبر الويب أو الاسترجاع
- ميزة الاستشهاد
- الوصول إلى الأدوات
- حالة الذاكرة
- حالة التعليمات المخصصة
- سياق الدردشة السابقة
- نوع الجلسة
- حالة التخصيص
- وضع الأمان أو السياسة، إذا كان معروفًا
- الميزات التجريبية
- تاريخ القياس والوقت UTC
- موجة القياس
- سجل تغييرات المنتج
- الثقة في الأدلة والسجلات

عندما يتغير أي من هذه الحقول بشكل جوهري:

- نسخة جديدة من منتج الذكاء الاصطناعي,

- خلية تجريبية جديدة,
- نسخة تحليلية جديدة

قد يكون مطلوبًا.

3. "نظام الذكاء الاصطناعي" ليس كائنًا واحدًا [K02; K03]

في هذه البروتوكول، "نظام الذكاء الاصطناعي" هو مصطلح عام ومرتفع المستوى. يمكن أن ينشأ استجابة مستخدم حقيقية من خلال الجمع بين عدة طبقات. سلسلة المرشح:

```
R= F(M, O, K, Q, T, P, S, U, C, L, t)
```

هنا:

- M: نموذج أساسي أو إرشادي
- O: طبقة تنسيق المنتج
- K: طبقة استرجاع أو مصدر المعرفة
- Q: إعادة الترتيب واختيار المصدر
- T: الأدوات
- P: طبقة السياسة والأمان
- S: الجلسة، الذاكرة، والتخصيص
- U: حساب المستخدم والخطة
- C: الدولة والمنطقة
- L: اللغة
- t: الوقت

الاستجابة التي يراها المستخدم:

R

الاستجابة المرئية للمستخدم ليست بالضرورة نتيجة النموذج الأساسي وحده. لذلك، فإن الكائن الرئيسي الموجه للمستهلك الذي يتم قياسه بواسطة GEO-1000 هو:

```
سلوك المنتج المقدم للمستخدم، وليس تجريد النموذج.
```

يمكن أيضًا إجراء بحث في واجهة برمجة التطبيقات أو النموذج بشكل مُسيطر عليه. هذه ليست نفس النتيجة التي تظهر في لوحة المستهلك الرئيسية.

4. سبع طبقات لهوية نظام الذكاء الاصطناعي

يجب تعريف كل نظام ذكاء اصطناعي مُدقق عبر سبع طبقات، بقدر ما تكون كل طبقة ذات صلة.

4.1. طبقة المزود

المنظمة التي تنشر أو تدير المنتج. اسم المزود:

- ليس هو نفسه

معرّف المنتج،

- معرّف النموذج،

- واجهة المستخدم

يمكن للمزود تقديم أكثر من منتج واحد. داخل نفس المنتج، يمكن استخدام أكثر من نموذج أو نظام تنظيم.

4.2. طبقة المنتج

المنتج التجاري أو الاستهلاكي أو المؤسسي أو التقني الذي يصل إليه المستخدم. أنواع أمثلة:

- منتج الدردشة العامة
- منتج الإجابة بمساعدة البحث
- مساحة عمل المؤسسات
- المنتج التعليمي
- المساعد المدمج
- واجهة برمجة التطبيقات للمطورين
- منتج مخصص قائم على المعرفة

اسم المنتج وحده لا يصف التكوين الكامل.

4.3. طبقة واجهة المستخدم

إنها الواجهة التي يتفاعل من خلالها المستخدم مع المنتج. ويب الجوال سطح المكتب امتداد المتصفح تكامل نظام التشغيل واجهة مدمجة ضمن المنصة

واجهة برمجة التطبيقات (API)

بوابة المؤسسة واجهات مختلفة لنفس المنتج:

- ميزات مختلفة
- نموذج مختلف،
- تمثيل مصدر مختلف،
- ذاكرة مختلفة

قد يتم استخدامها.

4.4. طبقة النموذج أو محرك الاستدلال

إنه النموذج أو عائلة النماذج أو نظام التوجيه المستخدم في توليد الاستجابة. معرف النموذج:

- مفتوح،
- مرئي جزئياً،
- غير معروف،
- ديناميكي

من الممكن. يمكن للمنتج توجيه استفسار المستخدم إلى نماذج مختلفة. في هذه الحالة، لا يمكن القول: 'تم قياس نموذج ثابت واحد.'

4.5. طبقة التنسيق والمعلومات

قد تتضمن ما يلي:

- الاسترجاع
- البحث المباشر على الويب
- البحث الداخلي

- تصنيف المصادر
- إعادة التصنيف
- اختيار الاستشهاد
- استدعاء مركبة
- إعادة صياغة الاستعلام
- تمزيق المستندات
- دمج الإجابات

هذه الطبقة لها أهمية أساسية لتمثيل GEO. لأنها يمكن أن تؤثر على أي سجل:

- يتم الوصول إليه،
- يتم اختياره،
- يتم تضمينه في الاستجابة،
- يتم عرضه كمرجع

يمكن التأثير عليه.

4.6. طبقة السياسة والتخصيص

قد تتضمن ما يلي:

- قواعد الأمان
- تعليمات النظام
- سياسة المنتج
- الذاكرة
- تعليمات خاصة
- تاريخ المستخدم
- القواعد الإقليمية
- العمر أو فئة الحساب
- إعدادات مسؤول الشركة

تؤثر هذه الطبقة على:

- نطاق الاستجابة،
- نغمتها،
- رفضها،
- توصيتها

يمكن أن تتغير.

4.7. طبقة التوزيع والزمن

يشير إلى أي مما يلي يكون المنتج خاضعًا له في لحظة معينة:

- ذات معنى مع
- ميزة تجريبية،
- التوزيع الإقليمي،
- قطاع المستخدمين،

- اختبار A/B

يمكن أن يكون لاسم المنتج نفسه شروط توزيع مختلفة لمستخدمين مختلفين في نفس اليوم.

5. مثال على منتج ذكاء اصطناعي

وحدة النظام الأساسية داخل GEO-1000:

مثال منتج ذكاء اصطناعي

المصطلح التشغيلي القياسي:

مثال منتج ذكاء اصطناعي

يتم تعريف مثال منتج ذكاء اصطناعي على أنه:

```
A_i= (v, p, s, r, m, k, t, x, u, c, l, ω)
```

يمكن تمثيله كالتالي. هنا:

- v: المزود
- p: المنتج
- s: واجهة المستخدم
- r: خطة أو فئة الحساب
- m: معرف النموذج أو العلامة
- k: الاسترجاع وهيكلة المعلومات
- t: تكوين الأداة
- x: الذاكرة، التخصيص، وحالة الجلسة
- u: فئة إعداد المستخدم أو المؤسسة
- c: الدولة والمنطقة
- l: اللغة
- ω: الوقت، النشر، وحالة الميزات التجريبية

ملاحظة: يجب ربط Oj بعينة منتج Ai محددة.

6. الفرق بين عينة المنتج واسم المنتج

يمكن تثبيت اسم المنتج. قد تختلف عينة المنتج. مثال: دع اسم منتج Orion AI يبقى بدون تغيير. ومع ذلك، هذان القياسان هما عينتان مختلفتان للمنتج:

عينة المنتج 1

خطة Web Free الذاكرة مغلقة الوصول إلى الويب مغلق محادثة جديدة تركية تركيا Wave 1

عينة المنتج 2

خطة Mobile Paid الذاكرة مفعلة الوصول إلى الويب مفعل سياق المحادثة السابقة متاح تركية تركيا Wave 1
يمكن أن توجد هذه تحت نفس العلامة التجارية. إنهما ليسا نفس الخلية التجريبية.

7. مستويات الثقة في هوية الطراز

لا يمكن دائمًا معرفة اسم الطراز بنفس درجة اليقين. لذلك، يمكن تسجيل هوية الطراز بالمستويات التالية.

MI-0— نموذج غير معروف

غير معروف أي طراز أو عائلة طراز يستخدمها المنتج. النتيجة تُنسب فقط إلى واجهة المنتج.

MI-1— الملصق المعروض للمستخدم

يكون ملصق الطراز مرئيًا في واجهة المستخدم. يتم تسجيل الملصق. لا يُفترض أنه يتحقق بالكامل من نشر الخلفية.

MI-2— عائلة الطراز الموثقة من قبل المزود

توصف وثائق المنتج الخاصة بالمزود عائلة نموذج محددة أو علاقة المنتج-النموذج. قد يظل من غير المعروف ما إذا كانت الملاحظة الفردية موجهة بدقة إلى ذلك النموذج.

MI-3— النشر المؤكد للجلسة

تتحقق الجلسة من توزيع محدد لواجهة برمجة التطبيقات أو بيانات المنتج الموثوقة.

MI-4— نقطة النهاية أو النشر المحدد بالإصدار

يمكن تسجيل إصدار محدد، أو نقطة نهاية، أو معرف نشر. إنه أقوى للمقارنة الخاضعة للرقابة.

MI-5— نموذج مجمد قابل لإعادة الإنتاج

هناك نموذج ثابت وقابل لإعادة الإنتاج أو هيكلية للنشر. قد يكون نادراً في المنتجات الاستهلاكية الحية. إذا كانت ثقة معرف النموذج منخفضة، يجب نسب النتيجة:

إلى سلوك المنتج

لا يمكن نسبها بشكل قاطع إلى نموذج محدد.

8. حد اسم النموذج الظاهر

رؤية اسم النموذج في واجهة المستخدم أمر ذو قيمة. ومع ذلك، قد تظل الأسئلة التالية مفتوحة: هل يتم توجيه جميع الاستفسارات إلى نفس النموذج؟ هل يتغير النموذج حسب صعوبة الاستفسار؟ هل تستخدم الأمان أو الاسترجاع نموذجًا منفصلًا؟ هل يطبق المزود نفس التصنيف على التوزيعات المختلفة؟ هل يشير تصنيف النموذج فقط إلى عائلة المنتج؟ هل المستخدم جزء من تجربة تجريبية؟ لذلك، قد يكون التقرير الصحيح: "تم عرض تصنيف النموذج X في واجهة المستخدم." بدون دليل أقوى، لا ينبغي استخدام الجملة التالية: "جميع الاستجابات يتم إنشاؤها بالتأكيد بواسطة النموذج X وحده."

9. نوع المنتج

يجب تصنيف كل منتج ذكاء اصطناعي بنوع أساسي وأنواع ثانوية.

PT-1— منتج المحادثة العامة

إنه منتج للمحادثة العامة والإجابة على الأسئلة.

—PT-2 منتج قائم على البحث أو الاسترجاع

يُعد البحث أو الاسترجاع مكونًا أساسيًا في توليد الإجابات.

—PT-3 منتج معرفة الشركات

يمكنه استخدام المستندات الداخلية أو مساحات العمل أو مصادر البيانات الخاصة.

—PT-4 مساعد مدمج

إنه مدمج في منتج أو منصة أخرى.

—PT-5 منتج للمطورين أو API

يستخدمه المطور أو التطبيق بدلاً من المستخدم النهائي.

—PT-6 منتج متخصص في مجال معين

تم تصميمه لمهنة أو صناعة محددة.

—PT-7 منتج متعدد الوسائط

تشمل المدخلات غير النصية، مثل المدخلات البصرية أو الصوتية أو المستندية، جزءًا من التجربة الرئيسية. يمكن أن يحمل المنتج أكثر من نوع واحد. يجب إجراء المقارنة ضمن نفس النوع أو ضمن نطاق صريح.

10. سجل واجهة المستخدم

يجب تسجيل السطوح المختلفة لنفس المنتج بشكل منفصل. مناطق الواجهة:

- معرّف الواجهة
- الويب / الهاتف المحمول / سطح المكتب / واجهة برمجة التطبيقات
- نسخة التطبيق
- نظام التشغيل، إن وجد
- المتصفح، إذا كان ذلك ممكنًا
- لغة المنتج
- اختيار الطراز المرئي
- توفر الأدوات
- عرض الاستشهاد
- ميزة تحميل الملفات
- إعدادات الذاكرة
- إعدادات إدارة الشركة
- الميزات التجريبية
- وقت القياس

إذا وُجد أن واجهتي الويب والجوال متشابهتان جوهريًا، يمكن دمج النتائج. لا يجب افتراض هذه المساواة؛ يجب اختبارها.

11. خطة وفئة الحساب

قد يحتوي نوع الخطة على ميزات يمكن أن تؤثر على استجابة المستخدم. حالات الخطة:

- مجاني
- فرد مدفوع
- مهني
- شركة
- التعليم
- مطور
- تجريبي
- دعوة أو وصول محدود
- غير معروف

قد تختلف الخطط في المجالات التالية:

- النموذج
- حد الاستخدام
- الوصول إلى الويب
- الذاكرة
- الأدوات
- الاستشهاد بالمصدر
- طول الاستجابة
- السرعة أو الأولوية
- بيانات المؤسسات
- إعدادات الأمان والاحتفاظ

إذا لم تتوفر معلومات عن الخطة:

PLAN_UNKNOWN

يجب استخدام الحالة. لا يمكن دمج النتائج المجانية والمدفوعة بصمت كسلوك منتج واحد.

12. الاسترجاع والوصول إلى الويب

يمكن توليد استجابة الذكاء الاصطناعي في واحدة من ثلاث حالات معرفية أساسية.

KG-0— وضع المعرفة غير المحدد

لا يمكن تحديد حالة الاسترجاع أو استخدام الويب.

KG-1— استجابة داخلية للنموذج

استجابة معروفة أو مفترض بقوة أنها لا تستخدم معلومات حية أو خارجية.

KG-2— استجابة مدعومة بالاسترجاع

قد يكون الاسترجاع قد تم استخدامه من مصادر مفهرسة مسبقًا أو من داخل المنتج.

KG-3— استجابة مباشرة عبر الويب أو بمساعدة البحث

تم استخدام الوصول إلى الويب/البحث في الوقت الفعلي أو القريب من وقت القياس.

KG-4— استجابة قائمة على المعرفة الخاصة

تم استخدام مصدر بيانات خاص قدمته المؤسسة أو المستخدم.

KG-5— استجابة هجينة أو متعددة المصادر

تم استخدام طبقات متعددة من المعلومات معًا. قد تنتج هذه الأنماط نتائج مختلفة في نفس المُدخل. يجب تمييزها بوضوح عند المقارنة.

13. الوصول للاسترجاع ليس هو نفسه استخدام الاسترجاع

قد يحتوي المنتج على ميزة البحث على الويب. ومع ذلك، قد لا يكون قد تم استخدامها في الإجابة المحددة. يجب تسجيل ثلاثة حقول بشكل منفصل: هل كانت الميزة متاحة؟ هل كانت الميزة مفعلة أو قابلة للتحديد؟ هل هناك دليل على أنه تم استخدامها في الإجابة المحددة؟ أمثلة على الحالات:

متاحة_لم تُستخدم

AVAILABLE_USAGE_UNKNOWN

مستخدمة_مؤكد

غير متاحة

معطلة

مجهول

وجود رابط مصدر بحد ذاته يمكن أن يكون مؤشراً قوياً على الاستخدام. عدم وجود رابط مصدر لا يثبت بالضرورة أن الاسترجاع لم يُستخدم.

14. طبقة الاقتباس

استخدام المصدر ليس هو نفسه عرض الاستشهاد. قد:

- يستخدم المنتج مصدرًا لكنه قد لا يعرضه،
- قد يعرض المنتج استشهادًا لكن الاستشهاد قد لا يدعم الادعاء،
- قد يربط المنتج نفس المصدر بجمل مختلفة.

يمكن للمستخدم إخفاء الاقتباس وفقًا لـ واجهة المستخدم. يجب أن يحتوي سجل الاقتباس على الحقول التالية: هل ميزة الاقتباس متاحة؟ هل تم عرض الاقتباس في الاستجابة المحددة؟ عدد الاقتباسات أسماء حقول المصدر تطابق الاقتباس—الادعاء الرؤية على واجهة المستخدم إمكانية توسيع الاقتباس إمكانية الوصول إلى المصدر وقت إنشاء الاقتباس المنتجات التي تعرض الاقتباسات والتي لا تعرضها لا يمكن مقارنتها مباشرة من حيث "عدد المصادر". أولاً، يجب شرح سلوك المنتج واختلافات الواجهة.

15. الوصول إلى الأدوات

قد يصل منتج الذكاء الاصطناعي إلى الأدوات التالية:

- بحث الويب
- المتصفح
- الآلة الحاسبة
- تنفيذ الكود
- قراءة الملفات
- خريطة
- تقويم
- قاعدة المعرفة المؤسسية
- كتالوج المنتجات
- واجهات برمجة التطبيقات الأخرى

قد يؤثر الوصول إلى الأدوات على ردك:

- مدى حداثته,
- الأدلة.
- نطاق الاستجابة،
- الدقة

سجل الأدوات:

- ما إذا كانت الأداة متاحة,
- ما إذا كانت نشطة,
- ما إذا تم استخدامها في الرد,
- ما إذا تم الاحتفاظ بنتيجة الأداة

يجب عرضها.

16. الذاكرة، التعليمات الخاصة، والتخصيص الشخصي

قد تتأثر الاستجابة المرئية للمستخدم بالحقول التالية:

- الذاكرة الدائمة
- سجل الدردشة
- تعليمات خاصة
- اسم المستخدم أو الملف الشخصي
- المحادثات السابقة مع العلامة التجارية
- التفضيلات
- الدولة والموقع
- إعدادات مساحة العمل المؤسسية
- التقسيم الذي يقوم به المزود

يجب تسجيل هذه الحقول في ثلاث طبقات منفصلة.

16.1. التخصيص الدائم

المعلومات المحمية عبر الحسابات أو الجلسات.

16.2. سياق المحادثة

الرسائل السابقة ضمن الدردشة الحالية.

16.3. النظام أو الإعداد التنظيمي

مسؤول أو تعليمات المنتج التي قد لا يراها المستخدم. المعلومات الكاملة قد لا تكون متاحة دائمًا. يجب تسجيل الحالات غير المعروفة صراحة.

17. أنواع الجلسات

حالات جلسة المرشح:

ST-1— محادثة قياسية جديدة

محادثة قياسية جديدة. يتم تسجيل الذاكرة أو التعليمات الخاصة بشكل منفصل.

ST-2— محادثة مؤقتة أو زائلة

جلسة مؤقتة يُقصد فصلها عن السجل الدائم. يجب التحقق من الميزات الفعلية في سجل المنتج.

ST-3— محادثة قائمة

هناك سياق محادثة سابقة. قد يكون ذا قيمة للوحة الطبيعية. يتم الاحتفاظ به بشكل منفصل في اللوحة المتحكمة.

ST-4— جلسة مساحة عمل للمؤسسة أو مشتركة

قد تكون هناك منطقة معلومات للمؤسسة أو مشتركة.

ST-5— حالة الجلسة غير معروفة

لا يمكن تحديد حالة الجلسة بشكل موثوق. إذا كان من المقرر دمج أنواع جلسات مختلفة في نفس الخلية، يجب توضيح الأساس والمنهجية والأهمية.

18. سياسة المنتج والرفض

قد لا يستجيب منتج الذكاء الاصطناعي لسؤال محدد بسبب:

- أمني،
- القانون،
- سياسة المنتج،
- القانون الإقليمي،
- لا يعني أي ملاءمة من حيث العمر،
- إعداد المؤسسة

السبب. الرفض:

- نقص في النموذج الأساسي،
- سياسة المنتج،

- واجهة
- قد يكون بسبب حساب المستخدم

تسجل سجلات التحكم بالمستخدم الرئيسي GEO-1000 نتيجة الرفض التي تصل إلى المستخدم. ومع ذلك، إذا كان سبب الرفض معروفًا، يتم تصنيفه إضافيًا. تعتبر طبقة سياسة المنتج أيضًا جزءًا من تجربة المستخدم.

19. الميزات التجريبية وتوزيعات A/B

قد لا تقدم منتجات الذكاء الاصطناعي الحية نفس الميزات لجميع المستخدمين في نفس الوقت. المشاركون:

- واجهة مستخدم مختلفة
- اختلاف صياغة المُدخلات باختلاف النماذج،
- ميزة ويب مختلفة،
- صيغة استجابة مختلفة

قد يُرى. مؤشرات الحالة التجريبية:

- ملصق بيتا على الواجهة
- توثيق المزود
- اختلاف الميزة
- اختلاف واجهة المستخدم النظامي بين المستخدمين
- حدوث كسر في السلوك أثناء القياس

قد يحدث. إذا كانت الميزة التجريبية غير معروفة:

EXPERIMENTAL_STATE_UNKNOWN

يجب استخدامه. الاختلاف الفعلي داخل المنتج ليس عيبًا في البيانات. إنها ميزة توزيع للنظام. ومع ذلك، يجب أخذها في الاعتبار عند المقارنة.

20. سجل نظام الذكاء الاصطناعي

يجب إنشاء سجل نظام ذكاء اصطناعي لكل منتج يتم تدقيقه. يجب أن يحمل السجل ثلاث مستويات.

20.1. سجل مقدم الخدمة والمنتج الرئيسي

يشمل المناطق الدائمة نسبيًا:

- المزود
- اسم المنتج
- نوع المنتج
- واجهات المستخدم الرئيسية
- الدول المصرح لها رسميًا بالوصول
- فئات الخطة
- سجلات اللغات المدعومة
- الخصائص العامة

20.2. سجل منتج موجة القياس

لموجة محددة:

- إصدار المنتج،
- ميزات الخطة،
- تسميات الطراز،
- ميزات الويب والأدوات،
- سلوك الذاكرة،
- حالة الوصول،
- التحديثات المعروفة

يتم تسجيلها.

20.3. سجل نظام مستوى الملاحظة

تمت الملاحظة فعليًا لكل مشارك:

- واجهة
- تسمية الطراز،
- الخطة،
- استخدام المركبة،
- استشهاد،
- الجلسة،
- التخصيص،
- الوقت

يحمل الحقول. السجل الرئيسي وحده لا يمكنه إزالة الاختلافات على مستوى الملاحظة.

21. بصمة تكوين النظام

يمكن إنشاء بصمة تكوين النظام لتسهيل قابلية المقارنة لنموذج AI معين. الحقول المرشحة:

```
SCF= H(canonicalise(p, s, r, m, k, t, x, c, l, w))
```

هنا، H هي الدالة التي تنتج قيمة النزاهة من الحقول المرقمة بالإصدار. البصمة:

- ما إذا كانت ملاحظتان تنتميان إلى نفس التكوين،
- ما إذا كانت هناك تغييرات جوهرية في الموجة

يمكن أن تساعد في التعرف عليها. البصمة: لا تثبت النموذج الخلفي الحقيقي، ولا تجعل الطبقات التي لم يكشف عنها المزود مرئية. إنها تعزز فقط ما إذا كان التكوين المسجل قد تم تغييره لاحقًا.

22. ما هي القابلية للمقارنة؟

يُعتبر نتيجتان قابلتين للمقارنة عندما:

> يمكن تفسير الفرق بينهما بشكل كافٍ من خلال سلوك المنتج الذي يتم مراقبته بدلاً من الاختلافات غير ذات الصلة في تصميم القياس.

القابلية للمقارنة ليست مساواة مطلقة. منتجات الذكاء الاصطناعي المختلفة:

- النموذج،
- المزود،
- واجهة المستخدم.

ليس من الضروري أن تكون جميعها متطابقة. ومع ذلك، يجب أن تكون الشروط المادية ذات الصلة بالمقارنة:

- المعادلة،
- التصنيف الطبقي،
- التسجيل،
- النمذجة في التحليل

مطلوب.

23. يجب تحديد موضوع المقارنة مسبقًا

يمكن لنفس البيانات أن تجيب على أسئلة مقارنة مختلفة.

23.1. مقارنة منتجات المستخدم

"أي منتج استهلاكي يقدم تمثيلًا أكثر دقة تحت نفس ظروف المستخدم؟" الكائن الذي تتم مقارنته هو سلسلة المنتج بأكملها.

23.2. مقارنة النماذج

"أي نموذج يقدم استجابة أكثر دقة تحت نفس المُدخل الخاضع للرقابة والظروف الخالية من الأدوات؟" قد يكون الوصول إلى API أو النموذج الثابت مطلوبًا.

23.3. مقارنة الاسترجاع

"هل هناك فرق بين الوضع الذي يمكن الوصول إليه عبر الويب والوضع غير القابل للوصول لنفس المنتج؟" يتم الحفاظ على النموذج والظروف الأخرى ثابتة قدر الإمكان.

23.4. مقارنة الخطة

"هل يرى المستخدمون المجانيون والمدفوعون نفس تمثيل العلامة التجارية؟" الفرق في الخطة هو موضوع التدقيق.

23.5. مقارنة الواجهة

"هل ينتج المنتج على الويب والهاتف المحمول تمثيلات مختلفة تحت نفس الظروف؟"

23.6. مقارنة الوقت

«هل تغير تمثيل نفس المنتج عبر موجات مختلفة؟» يجب فصل تغييرات المنتج والتعداد السكاني. بدون تحديد سؤال المقارنة: «المنتج A أفضل من المنتج B.» هو واسع جدًا.

24. أبعاد المقارنة

عند مقارنة منتجين أو موجتين، يجب تقييم المجالات التالية:

- الكيان الذي تم تدقيقه
- إصدار حزمة الواقع
- معرّف الإصدار والإصدار

- اللغة والمنطقة
- البلد
- السكان المستهدفون
- مصدر العينة
- تركيب المشاركين
- نوع الجلسة
- الذاكرة
- تعليمات خاصة
- المحادثة السابقة
- خطة المنتج
- واجهة المستخدم
- تصنيف النموذج
- حالة الاسترجاع/الويب
- الأدوات
- ميزة الاستشهاد
- وقت القياس
- تحديثات المنتج
- قاعدة التحكيم
- حالات النتائج
- الوزن

لا تحتاج جميع الأبعاد إلى أن تكون متماثلة. الأبعاد المختلفة تحد من نطاق المقارنة.

25. مستويات القابلية للمقارنة

CP-0— غير قابلة للمقارنة

شروط قياس المادة غير معروفة أو غير متسقة بشكل كبير. لا يمكن إجراء مقارنة مباشرة للأداء.

CP-1— التواجد الوصفي المشترك

هناك نتائج من منتجين أو موجتين. ومع ذلك:

- السكان،
- مُدخل,
- الوقت،
- الجلسة

ليست متساوية. يمكن إجراء عرض على الجانب الوصفي فقط.

CP-2— متوافقة جزئيًا

بعض الحقول الأساسية متطابقة. الفروقات المادية مستمرة. يمكن إجراء مقارنة محدودة.

—CP-3 مقارنة المستخدم-الواجهة المتطابقة

نفس الكيان، نفس إصدار المُدخل، نفس اللغة/البلد، مجتمع مستخدم مشابه، نفس نافذة الوقت، ميزات المنتج المسجلة متاحة. نتائج المنتجات الاستهلاكية قابلة للمقارنة.

—CP-4 مقارنة متحكم بها ومتطابقة

بالإضافة إلى ذلك:

- الجلسة،
- الذاكرة،
- التخصيص،
- الاسترجاع،
- الأدوات،
- خطة

متطابقة بطريقة خاضعة للرقابة أو مفتوحة. يوفر مقارنة أقوى للمنتج أو الميزة.

—CP-5 مقارنة مكررة ومعايرة

المقارنة:

- في موجات متعددة،
- مع مستخدمين أو عينات مستقلة،
- مع نفس التقييم المعياري

تم التكرار؛ لقد تم معايرة التغيرات وتغييرات المنتج. هذه المستويات ليست "درجة الدقة العلمية". إنها تشير إلى الحد الذي يمكن لادعاء المقارنة تحمله.

26. ماذا تفعل في حالة CP-0؟

إذا لم يكن بالإمكان مقارنة منتجين، فإن النتائج لا يتم تجاهلها. يمكن الإبلاغ عنها بشكل منفصل:

- «هذه هي نتيجة نطاق الوصول الطبيعي للمنتج أ.»
- هذا هو نتيجة المنتج ب تحت مستخدمين مختلفين وظروف المنتج المختلفة.

ومع ذلك: لا يمكن القول، "أ أفضل من ب بمقدار 12 نقطة." إن عدم القابلية للمقارنة ليس فشلاً، بل هو حد من حدود الطريقة.

27. القابلية للمقارنة لنفس المزود

لا يمكن مقارنة النتائج من نفس المزود تلقائيًا. قد تكون هناك الاختلافات التالية:

- المنتج
- الخطة
- النموذج
- الوصول إلى الويب
- واجهة
- الذاكرة
- بيانات المؤسسات

- توزيع الدولة
- ميزة تجريبية
- الوقت

يجب أيضًا تقييم مقارنة نفس المزود وفقًا لمستوى CP.

28. القابلية للمقارنة لمزودين مختلفين

منتجات المستخدمين التي تنتمي إلى مزودين مختلفين:

- نفس الفئة المستهدفة,
- نفس المحفز,
- نفس الوقت،
- مماثل واجهة المستخدم
- تكوين المنتج واضح

يمكن مقارنته أدناه. كونك مزودًا مختلفًا لا يمنع المقارنة. ومع ذلك، فإن الاختلافات في الميزات تحدد ما تقيسه المقارنة. على سبيل المثال: يستخدم المنتج أ الويب المباشر. المنتج ب لا يستخدمه. يمكن أن يجيب الناتج على هذا السؤال: 'أي من سلسلتي المنتجين المقدمتين للمستخدم يوفر تمثيلًا أكثر دقة؟' ولا يجيب مباشرة على هذا السؤال: 'أي نموذج أساسي أكثر معرفة؟'

29. مقارنة المنتجات المدعومة بالويب وغير المدعمة به

يمكن إجراء مقارنتين منفصلتين.

29.1. مقارنة المنتجات الطبيعية

يتم قياس كل منتج في ظروف المستخدم الحقيقي الافتراضية. المنتج الذي يستخدم الويب يستخدم الويب. والذي لا يستخدمه لا يستخدمه. النتيجة: إنها مقارنة لتجربة المنتج الحقيقي للمستهلك.

29.2. مقارنة مطابقة للميزات

إذا أمكن، كلا المنتجين:

- تم قياسهما في حالة فتح الويب،
- وحالة إغلاق الويب

كل على حدة أيضاً. هذا يمكن أن يفصل تأثير الاسترجاع بشكل أفضل. قد لا يقدم كل منتج نفس الخيارات. الوضع المفقود:

تسجَّل كغير متاح

يجب أن يظل هذا الوضع صريحًا.

30. المنتجات التي تظهر ولا تظهر الاستشهادات

بالنسبة للمنتج الذي لا يملك ميزة الاستشهاد: من الممكن إعطاء "صفر استشهاد". ومع ذلك، يجب توضيح معناه: قد يكون المنتج لا يستشهد بالمصادر. ولا يمكن التأكد من أنه لا يستخدم أي مصادر. تغطية الاستشهاد ودقة التغطية هما نتائج منفصلة. يجب عمل مقارنة الاستشهاد فقط بين المنتجات التي تدعم الاستشهاد أو مع حالة ميزة منفصلة.

31. نطاق الوصول الطبيعي المقارنة

يتم قياس كل منتج ضمن عدد السكان الفعلي الذي يصل إليه. الميزة: يظهر التأثير العالمي الطبيعي للمنتج. القيد: عدد السكان مختلف، وقد تختلف توزيعات الدول واللغات. درجات نطاق الوصول الطبيعي: توضح ما يفعله المنتج ضمن السكان الذين يصل إليهم. لا يجب استخدامها بمفردها في تصنيف تفوق المنتج مباشرة.

32. مقارنة على أساس مشترك

يتم استخدام شريحة الوصول المشتركة للمنتجات التي تمت مقارنتها. الميزة:

- نفس الدولة،
- نفس اللغة،
- نفس حالة المستخدم

يوفر مقارنة أكثر عدلاً. القيد: يحافظ على الفارق الفعلي في الوصول العالمي للمنتجات، قد يكون الكون المشترك ضيقًا جدًا. يجب نشر درجة قاعدة المقارنة المشتركة مع درجة نطاق الوصول الطبيعي.

33. مقارنة المستخدمين المزدوجة

إذا قام نفس المستخدم باختبار منتجين أو أكثر من منتجات الذكاء الاصطناعي، يمكن التحكم في خصائص المستخدم. ومع ذلك:

- ترتيب المهام,
- استدعاء الإجابة,
- الألفة مع المنتج,
- الملف الشخصي الخاص بالمستخدمين الذين يمكنهم الوصول إلى جميع المنتجات

يمكن أن يخلق تحيزًا. في تصميم مزدوج:

- يجب أن يكون ترتيب المنتج عشوائيًا،
- يجب عدم مشاركة الردود السابقة

يجب الحفاظ على نفس تبعية المستخدم في التحليل.

34. مقارنة المستخدم المستقل

يتم اختبار كل منتج بواسطة مستخدمين مختلفين، ولكن يتم اختيارهم من نفس الإطار السكاني. الميزة: يتم تقليل تلوث المهمة، ويمكن اختيار مستخدمين للمنتج أكثر طبيعية. القيد: قد لا تكون مجموعات المستخدمين متساوية تمامًا. يلزم الوزن والتصنيف. يمكن فصل نتائج المستخدمين المزدوجين والمستقلين ولكنها تظل مكملة.

35. قابلية مقارنة الوقت

عند مقارنة نتائج اليوم لمنتج ذكاء اصطناعي مع النتائج السابقة، يجب فحص التغييرات التالية:

- النموذج
- إصدار المنتج
- الويب/البحث

- الخطة
- وصول الدولة
- دعم اللغة
- مُدخل
- واقع الكيان
- التحكيم
- إطار السكان
- الوزن

قد لا ينشأ الفرق الزمني فقط من منتج الذكاء الاصطناعي.

36. أنواع تغييرات المنتج

UC-1— التغيير الصغير الموثَّق

تغيير صغير كشف عنه المزود وتم تقييمه على أنه لا يغير جوهريًا موضوع القياس.

UC-2— التغيير المادي الموثَّق

تغيير جوهري موصوف في النموذج أو الاسترجاع أو الواجهة أو الميزة.

UC-3— تغيير السلوك الملحوظ

حتى إذا لم يكن هناك إعلان رسمي، يتم ملاحظة سلوك منظم أو تغيير في الواجهة.

UC-4— تغيير الوصول أو الخطة

تغييرات في الدولة أو الخطة أو وصول المستخدم.

UC-5— حدث التغيير غير معروف

هناك انقطاع في سلوك القياس. لم يمكن التحقق من السبب.

UC-6— استبدال المنتج أو التقاعد

انتهى منتج أو نموذج قديم وتم استبداله بهيكل آخر. نوع التغيير يحد من مقارنة السلاسل الزمنية.

37. تغيير المنتج ضمن موجة القياس

إذا حدث تغيير مادي خلال الموجة، تتوفر الخيارات التالية: يمكن إيقاف الموجة وإعادة تشغيلها. يمكن عمل موجات فرعية منفصلة قبل وبعد التغيير. يمكن نمذجة التغيير كعامل تجريبي. تُصنف الموجة على أنها حالة نظام مختلط. الهيكل التالي محظور: دمج جميع الردود كما لو كانت نتيجة منتج واحد ثابت على الرغم من معرفة التغيير.

38. يمكن أن يتغير السلوك دون تغيير المنتج

حتى لو لم يعلن المزود عن تحديث صريح، قد تتغير النتيجة. الأسباب المحتملة:

- تغيير في فهرس الاسترجاع
- تحديث الموارد

- ظهور محتوى جديد على الويب
- تغيير في إعادة الترتيب
- اختبار A/B
- التوزيع الإقليمي
- سياسة الأمان
- توجيه حركة المرور

لذلك، لا يمكن القول، "لم يقم المزود بالتحديث، النظام بالتأكيد كان هو نفسه." يجب أيضًا فحص أدلة القياس وتحليل السلوك.

39. المنتج الحي والنموذج المجمد

هناك كائنان بحثيان مختلفان.

39.1. تدقيق المنتج الحي

يقيس ما يراه المستخدمون الحقيقيون على واجهة المستهلك الحقيقي. الميزة:

- يعرض التجربة الحقيقية،
- المصدر الحقيقي،
- السياسة الحقيقية،
- الواجهة الحقيقية

القيود: أنها متغيرة، وقد يكون الاستنساخ الكامل صعبًا.

39.2. تدقيق النموذج المجمد

يتم إجراء تكرار مضبوط تحت إصدار أو نقطة نهاية ثابتة. الميزة:

- يوفر إمكانية إعادة الإنتاج،
- السيطرة التجريبية

القيود: قد لا تمثل المنتج الاستهلاكي الحقيقي بشكل كامل. الطبقة السكانية الرئيسية لـ GEO-1000 NOMOS تعتمد على المنتج المستخدم مباشرة من قبل المستخدمين. تدقيق النموذج المجمد هو عمل مخبري تكميلي.

40. عدم توفر المنتج والنتائج التقنية

يجب الاحتفاظ بالحالات التالية بشكل منفصل:

AR-1— المنتج غير متوفر

المنتج غير متوفر للبلد أو لمجموعة المستخدمين. هذا نتيجة لنطاق الوصول.

AR-2— الحساب أو الخطة غير مؤهل

الخطة أو حالة الحساب المعينة للمشارك غير مؤهلة. قد يكون هناك مشكلة في العينة أو تعيين المهام.

AR-3— فشل تقني مؤقت

حدث خطأ تقني مؤقت. يجب أن يكون بروتوكول إعادة المحاولة محددًا مسبقًا.

AR-4— حد معدل الاستخدام أو التقييم

تم الوصول إلى حصة الاستخدام أو حدود المعدل. قد يكون جزءًا من تجربة المستخدم الحقيقية.

AR-5— رفض النظام

تم رفض مهمة المنتج بسبب السياسة أو المحتوى. هذا نتيجة صحيحة للنظام.

AR-6— لا يوجد مخرج قابل للاستخدام

لم يتم توليد أي استجابة قابلة للاستخدام.

AR-7— فشل الوصول غير معروف

لماذا لم يكن بالإمكان تحديد ذلك. لا يمكن دمج هذه الحالات في نفس صندوق "عدم الرد".

41. قاعدة إعادة المحاولة التقنية

يمكن إعادة المحاولة في حالة حدوث خطأ تقني مؤقت أو انقطاع في الاتصال. ومع ذلك:

- إذا لم يكن الرد من الذكاء الاصطناعي مرغوبًا،
- حدث رفض،
- تم إنتاج معلومات غير صحيحة

لا يمكن إعادة المحاولة لهذه الأسباب. إعادة المحاولة التقنية:

- نفس المستخدم،
- نفس المنتج،
- نفس التكوين،
- يجب أن تتم في الفترة المحددة مسبقًا.

يجب أيضًا تسجيل أول حادث فني.

42. مصفوفة المقارنة

يمكن أن يكون سجل المقارنة الأساسي لمنتجين ما كما يلي:

البعد	المنتج أ	المنتج ب	الحالة
الكيان المستهدف	متطابق	متطابق	متوافق
نسخة المُدخل	1.0	1.0	متوافق
اللغة/المنطقة	tr-TR	tr-TR	متوافق
البلد	تركيا	تركيا	متوافق
نافذة الوقت	الموجة 1	الموجة 1	متوافق
لوحة المستخدم	السكان	السكان	متوافق
الخطة	مجاني	مدفوع	مختلف
الواجهة	الويب	الهاتف المحمول	مختلف
الوصول إلى الويب	مغلق	مفتوح	مختلف

البعد	المنتج أ	المنتج ب	الحالة
الذاكرة	مغلق	مفتوح	مختلف
معرّف النموذج	MI-1	MI-1	محدود
مستوى المقارنة			CP-2

يمكن عرض هذه النتائج جنبًا إلى جنب. ومع ذلك: "النموذج الأساسي من المنتج B أفضل." لا يمكن قوله. قد تنبع الفروقات من سلسلة المنتج بأكملها.

43. حالة مقارنة APPLE.COM الاصطناعية

عرض منهجية اصطناعية / المنتجات، الردود، والنتائج أدناه كلها خيالية تمامًا. ولا تمثل أداء شركة أبل الحقيقية أو أي مزود ذكاء اصطناعي حقيقي. لنأخذ ثلاثة أمثلة اصطناعية للمنتجات.

مثال المنتج A

المزود: أوريون المنتج: أوريون تشات الواجهة: ويب الخطة: مجانية تصنيف النموذج: أوريون ستاندرد ثقة النموذج: MI-1 الوصول إلى الويب: مغلق الذاكرة: مغلقة الجلسة: جديدة اللغة: التركية المنطقة: تركيا الموجة: W1 الرد: "آبل شركة أمريكية للتكنولوجيا تنتج الهواتف وأجهزة الكمبيوتر."

عينة المنتج ب

المزود: أوريون المنتج: أوريون تشات الواجهة: موبايل الخطة: مدفوعة تصنيف النموذج: أوريون برو ثقة النموذج: MI-1 الوصول إلى الويب: مفتوح الاستشهاد: مفتوح الذاكرة: مفتوحة المحادثة السابقة: عن آبل اللغة: التركية المنطقة: تركيا الموجة: W1 الرد: "آبل شركة تكنولوجيا عالمية تقدم الأجهزة والبرمجيات والخدمات الرقمية. يمكن العثور على المعلومات الحالية عن الشركة في هذه المصادر."

عينة المنتج ج

المزود: هيليوس المنتج: هيليوس البحث الواجهة: الويب الخطة: مجانية تصنيف النموذج: غير معروف ثقة النموذج: MI-0 الوصول إلى الويب: مفتوح الاستشهاد: مفتوح الذاكرة: لا شيء الجلسة: جديدة اللغة: التركية المنطقة: تركيا الموجة: W1 الاستجابة: "آبل هي شركة تكنولوجيا أمريكية تطور أجهزة استهلاكية وأنظمة تشغيل وخدمات رقمية."

43.1. مقارنة بين A و B

هم نفس المزود. ومع ذلك:

- الخطة،
- الواجهة،
- تسمية الطراز،
- الويب،
- الذاكرة،
- المحادثة السابقة

مختلف. يمكن أن يكون مستوى المقارنة CP-1 أو CP-2. لا يمكن عزو الفرق في الاستجابة فقط إلى اختلاف النموذج الأساسي.

43.2. مقارنة بين A و C

هم مزودون مختلفون. ومع ذلك:

- جلسة جديدة،
- الخطة المجانية،
- واجهة الويب،
- نفس اللغة/البلد،
- نفس الموجة

من حيث بعض الشروط تتطابق. الوصول إلى الويب مختلف. المقارنة: يمكن القيام بها كنتيجة طبيعية لمنتج للمستهلك. لا يمكن القيام بها كقدرة أساسية للنموذج.

43.3. مقارنة بين B و C

كلا المنتجين يقدمان الوصول إلى الويب وتمكين الاستشهادات. ومع ذلك، يمتلك B ذاكرة ومحادثات سابقة، بينما C لا يمتلك ذلك. تجربة المستخدم للمنتجات قابلة للمقارنة. ليست مقارنة منتج نظيفة ومتحكم فيها.

43.4. تكرار محسوب

في المنتج B:

- تم إيقاف تشغيل الذاكرة،
- تم بدء محادثة جديدة،
- تم إيقاف التعليمات الخاصة،

الويب يظل مفعلًا. يتم قياس C تحت نفس الظروف. مستوى المقارنة: CP-3، أو CP-4 مع وجود تحكم إضافي كافٍ. يوضح هذا المثال:

> المزود نفسه لا يقدم مقارنة قريبة؛ قد تكون مقدمو الخدمات المختلفون المسجلون جيدًا أكثر قابلية للمقارنة.

44. حالة تغيير المنتج الاصطناعي

مثال اصطناعي

يستهدف W1 Orion Chat موجة مراقبات 1,000. في أول مراقبات 480:

- نسخة المنتج: 6.1
- ميزة الويب: يدوي
- النسبة: محدودة

في الملاحظات التالية 520:

- نسخة المنتج: 6.2
- الميزة على الويب: تلقائية
- النسبة: موسعة

ليكن كذلك. النتيجة الفردية: إذا تم حسابها على أنها "درجة W1 Orion Chat"، يتم خلط حالتين مختلفتين للنظام. الخيارات الصحيحة:

الخيار 1

AW1: أول 480 BW1: التالي 520 يتم الإبلاغ عنها بشكل منفصل.

الخيار 2

الموجة غير مستمرة. تُؤخذ ملاحظات جديدة 1,000 في الإصدار 6.2.

الخيار 3

يتم الاحتفاظ بالإصدار كعامل تجريبي صريح في التحليل الإحصائي. الخيار غير الصحيح: يتم إزالة التحديث من التقرير وتُعرض جميع الملاحظات كمنتج ثابت واحد.

45. الحد الأدنى لسجل نظام الذكاء الاصطناعي

يجب أن يحتوي كل ملاحظة على الحقول التالية على الأقل:

- معرف نظام الذكاء الاصطناعي NOMOS
- المزود
- المنتج
- الواجهة
- الخطة
- نوع المنتج
- العلامة النموذجية الظاهرة
- مستوى ثقة النموذج
- البلد
- الموقع
- اللغة
- نوع الجلسة
- حالة الذاكرة
- حالة التعليمات المخصصة
- حالة المحادثة السابقة
- ميزة الاسترجاع/الويب
- حالة استخدام الاسترجاع
- حالة الاقتباس
- حالة الأداة
- حالة الميزة التجريبية
- الوقت UTC
- الموجة
- بصمة النظام
- أحداث التغيير
- مصدر الدليل
- مالك السجل

إذا كان حقل المادة غير معروف، لا يمكن تركه فارغًا. يجب استخدام حالة غير معروف.

46. بطاقة المقارنة الدنيا

عند مقارنة منتجين أو موجتين، يجب أن يظهر السجل العام:

- سؤال مقارنة
- مستوى المقارنة
- السكان المستهدفون الشائعون
- نطاق الوصول الطبيعي الاختلافات
- مساواة المستخدمين
- اللغة والبلد
- وقت القياس
- المنتج، الخطة، والواجهة
- ثقة هوية النموذج
- الفرق بين الويب والاسترجاع
- الاختلاف في الذاكرة والتخصيص
- الاختلاف في الأداة
- الاختلاف في العينة
- طريقة التقييم
- قيود المادة
- المقاييس القابلة للمقارنة مباشرة
- المجالات غير القابلة للمقارنة

لا ينبغي نشر رقم واحد أو ترتيب بدون هذه البطاقة.

47. قاعدة القابلية للمقارنة

يجب على المدقق الإجابة على هذه الأسئلة الأربعة: هل تم سؤال الشيء نفسه؟ هل تم قياس نفس الكون البشري؟ هل تم تطبيق نفس ظروف المستخدم؟ هل تم استخدام نفس الوقت ونظام التقييم؟ كل من هذه الأسئلة:

- تمامًا،
- جزئيًا،
- لا,
- غير معروف

يتم تصنيفه على هذا النحو. يتم تحديد مستوى المقارنة وفقًا لذلك.

48. الأحكام الإلزامية المعيارية

CH09-N01

يجب ربط كل ملاحظة GEO-1000 بتسجيل مسجل نظام الذكاء الاصطناعي النسخي.

CH09-N02

لا يمكن استخدام اسم المزود كهوية كاملة لمنتج الذكاء الاصطناعي المقاس.

CH09-N03

يجب فصل المنتج واجهة المستخدم والخطة والطراز والاسترجاع والأداة والجلسة وطبقات التخصيص عن بعضها البعض.

CH09-N04

لا يمكن تقديم تسمية الطراز على واجهة المستخدم كهوية نموذج خلفية مؤكدة.

CH09-N05

يجب تسجيل مستوى الثقة في هوية الطراز من MI-0 إلى MI-5 أو من خلال حالة معرّفة علنًا مكافئة.

CH09-N06

إذا لم يمكن التحقق من هوية الطراز، يجب نسب النتيجة فقط إلى المنتج و واجهة المستخدم.

CH09-N07

المنتجات التي تنتمي إلى نفس المزود لا تُعتبر تلقائيًا نفس الخلية التجريبية.

CH09-N08

قد تكون المنتجات التي تنتمي إلى مزودين مختلفين قابلة للمقارنة عندما تتطابق المواد المستخدمة وشروط التجربة.

CH09-N09

لا يمكن دمج واجهات منتجات الويب، والهاتف المحمول، وواجهة المكتب، والمضمنة، وواجهة برمجة التطبيقات دون وصف.

CH09-N10

يجب تسجيل الخطط المجانية، والمدفوعة، والمؤسسية، والتعليمية، وخطط المطورين كمصطلحات منتج منفصلة.

CH09-N11

إذا كانت الخطة غير معروفة، يجب أن تحمل النتيجة حالة PLAN_UNKNOWN.

CH09-N12

لا يمكن اعتبار توفر ميزة الاسترجاع أو الويب دليلًا على أنه تم استخدامها في إجابة معينة.

CH09-N13

يجب تسجيل الوصول إلى الاسترجاع، واستخدام الاسترجاع، والنسب بمحددات منفصلة.

CH09-N14

لا يمكن اعتبار المنتج الذي لا يظهر نسبًا على أنه لم يستخدم أي موارد تلقائيًا.

CH09-N15

يجب فحص الاستشهادات الخاصة بالمنتج الذي يتم الاستشهاد به بشكل منفصل من حيث دعم الادعاء.

CH09-N16

يجب تسجيل الوصول إلى الأدوات واستخدام الأدوات في إجابات محددة بشكل منفصل.

CH09-N17

يجب تسجيل الذاكرة والتعليمات الخاصة والدردشات السابقة والإعدادات الخاصة بالشركة بشكل منفصل قدر الإمكان.

CH09-N18

لا يمكن افتراض أن حالة التخصيص غير المعروفة تشبه جلسة نظيفة.

CH09-N19

لا يمكن دمج الجلسات الجديدة، أو المؤقتة، أو الحالية، أو جلسات بيئة العمل المؤسسية دون تفسير.

CH09-N20

يجب أن تحمل نتائج عينة المستخدمين في ظروف الاستخدام الطبيعية و عينة اختبار في ظروف منضبطة حالة المنتج بشكل منفصل.

CH09-N21

إذا كانت ميزة تجريبية أو توزيع A/B في المنتج جوهرية، فيجب الاحتفاظ بها في سجل الملاحظة.

CH09-N22

إذا حدث تغيير جوهري في المنتج خلال الموجة، فيجب تقسيم الموجة إلى نسخ فرعية منفصلة، أو إعادة تشغيلها، أو يجب نمذجة التغيير بشكل صريح في التحليل.

CH09-N23

لا يمكن تخزين تغيير معروف في المنتج كما لو أن جميع الملاحظات جاءت من نظام ثابت واحد.

CH09-N24

لا يمكن اعتبار عدم قيام المزود بشرح تحديث ما دليلاً قاطعًا على أن المنتج لم يتغير.

CH09-N25

لا يمكن تقديم نتيجة منتج تجاري حي كما لو كانت نتيجة نموذج قاعدة بيانات مجمدة.

CH09-N26

لا يمكن تقديم نتيجة نموذج مجمد في المختبر على أنها أداء منتج يستخدمه مستخدم حقيقي.

CH09-N27

يجب تسجيل غياب المنتج، أو خطأ تقني، أو حد الاستخدام، أو رفض النظام، أو عدم وجود نتيجة بحالات منفصلة.

CH09-N28

لا يمكن أن يكون الرد الخاطئ أو السلبي للذكاء الاصطناعي سببًا لإعادة المحاولة التقنية.

CH09-N29

يجب تحديد موضوع المقارنة قبل جمع البيانات.

CH09-N30

يجب أن يكون مستوى قابلية المقارنة بين منتجين مرئيًا في النتائج العامة.

CH09-N31

لا يمكن إنشاء تصنيف محدد للمنتج من نتائج CP-0 أو CP-1.

CH09-N32

استخدام صياغة المُدخل نفسها بمفردها لا يخلق تجربة قابلة للمقارنة.

CH09-N33

في المقارنات، يجب فحص الكيان المستهدف، وحزمة الحقيقة، وصياغة المُدخل، واللغة، والبلد، والوقت، والسكان، والجلسة، والمنتج، وطريقة التقييم معًا.

CH09-N34

لا يمكن تصنيف نتائج نطاق الوصول الطبيعي مباشرة بدون شرح لأنها تنتمي إلى جماهير مستهدفة مختلفة.

CH09-N35

لا يمكن عرض نتيجة قاعدة المقارنة المشتركة كنتيجة الوصول العالمية الفعلية للمنتج.

CH09-N36

يمكن إجراء مقارنة بين المنتجات المدعومة بالويب والخالية من الويب كمقارنة منتجات من قبل المستخدم؛ لا يمكن تقديمها كأفضلية لنموذج أساسي.

CH09-N37

في المقارنات المزدوجة التي يتم إجراؤها مع نفس المستخدم، يجب الحفاظ على ترتيب المنتج واعتماد المستخدم.

CH09-N38

في المقارنات التي يتم إجراؤها مع مستخدمين مختلفين، يجب أن يكون تكوين المستخدم وفروق الوزن مرئيين.

CH09-N39

يجب تخزين بصمة إصدار المنتج أو تكوينه في كل موجة قياس.

CH09-N40

يجب أن يكون هناك مالك بشري أو مؤسسي مسؤول عن سجل النظام وقرار القابلية للمقارنة.

49. أشكال الفشل

CH09-F01— النظر في المزود كذكاء اصطناعي واحد

جميع المنتجات والنماذج والواجهات الخاصة بالمزود معروضة كنظام واحد.

CH09-F02— عد سلوك المنتج حسب اسم النموذج

يتم تجاهل الاسترجاع والأمان والواجهة والتخصيص.

CH09-F03— اعتبر بالتأكيد الملصق المرئي للنموذج كخلفية

يتم تجاهل استخدام التوجيه أو النماذج الديناميكية.

CH09-F04— اعتبار نفس المزود كمقارن

يتم إخفاء الفروق في الخطة، الواجهة، والميزات.

CH09-F05— اعتبار مزودين مختلفين كغير قابلين للمقارنة

يتم رفض مقارنة المنتجات الممكنة تحت شروط المستخدم المزدوج.

CH09-F06— دمج الويب والهاتف المحمول

تختفي الاختلافات في الواجهة والميزات.

CH09-F07— الجمع بين الخطط المجانية والمدفوعة

تصبح اختلافات الجودة أو الميزات بسبب الخطة غير مرئية.

CH09-F08— اعتبار المنتجات المؤسسية والاستهلاكية متساوية

يتم إخفاء المعلومات الخاصة وإعدادات المسؤول.

CH09-F09— التعامل مع نتيجة واجهة برمجة التطبيقات كمنتج استهلاكي

تم إلغاء الاختلافات بين التنسيق و واجهة المستخدم.

CH09-F10— إذا وجدت الاسترجاعات، اعتبرها مستخدمة

لا يمكن التحقق من مسار المعرفة الحقيقي للإجابة المحددة.

CH09-F11— إذا لم توجد اقتباسات، اعتبر أنه لا يوجد مصدر

يتم تجاهل احتمال وجود استرجاعات مخفية أو غير مرئية.
CH09-F12— إذا وُجد اقتباس، احسبه كمصدر صحيح
لم يتم فحص مطابقة الاقتباس مع الادعاء.
CH09-F13— اعتبر الوصول إلى الأداة استخدامًا للأداة
من غير المعروف أي أداة تم استخدامها في الإجابة المحددة.
CH09-F14— اعتبر الذاكرة كجلسة نظيفة
يبقى تأثير التخصيص غير مرئي.
CH09-F15— اعتبر المحادثة الجديدة غير مخصصة
يتم تجاهل إعدادات الحساب المستمرة والتعليمات الخاصة.
CH09-F16— عرض المحادثة الحالية كجلسة جديدة
يتم إخفاء السياق السابق.
CH09-F17— احتساب ميزة تجريبية كخطأ في البيانات
يتم إعلان التغير الفعلي في المنتج كملاحظة غير صالحة.
CH09-F18— إخفاء ميزة تجريبية
يتم دمج المستخدمين في توزيعات مختلفة كمنتج ثابت واحد.
CH09-F19— موجة واحدة بعد التحديث
تغيير المنتج المادي يتم نشره كنتيجة لنظام ثابت واحد.
CH09-F20— تجاهل التغييرات إذا لم يكن هناك إعلان
سلوك الانكسار الملحوظ لم يتم فحصه.
CH09-F21— احتساب المنتج الحي كنموذج مجمد
يتم الادعاء بإمكانية التكرار الكامل.
CH09-F22— اعتبار نموذج مجمد كمنتج مستخدم حقيقي
سلسلة المنتج واختلافات الواجهة مخفية.
CH09-F23— احتساب غياب المنتج كخطأ
نطاق الوصول يختلط مع دقة التمثيل.
CH09-F24— رفض خطأ تقني
يتم تصنيف مشكلة الاتصال أو الخدمة كنتيجة للسياسة.
CH09-F25— احتساب خطأ تقني كمرفوض وإعادة المحاولة
يتم إزالة نتيجة المستخدم الفعلية من المقام.
CH09-F26— إخفاء حد الاستخدام
تصبح التجربة الحقيقية للخطة المجانية غير مرئية.
CH09-F27— اعتبار نفس الطلب كأنّه نفس التجربة
يتم تجاهل الاختلافات في المنتج، والخطة، والجلسة، والويب، والوقت.
CH09-F28— اعتبار نتيجة المنتج كأفضلية للنموذج
يُعزى الفرق عبر سلسلة النظام بأكملها إلى النموذج الأساسي.

CH09-F29— الترتيب المباشر لدرجات نطاق الوصول الطبيعي
يُعتبر كون السكان المختلفون واحدًا.
CH09-F30— اعتبار قاعدة المقارنة المشتركة كنطاق عالمي
يتم عرض الكون المشترك الضيق كالوصول الكامل للمنتج.
CH09-F31— الادعاء بالريادة من نتيجة CP-1
تُستمد تفوق المنتج فقط من النتائج غير المتوافقة التي توضع جنبًا إلى جنب.
CH09-F32— الحفاظ على مستوى المقارنة
العامة ترى الترتيب فقط.
CH09-F33— تجاهل اعتماد المستخدم المزدوج
تعتبر ردود نفس الشخص على المنتجات مستقلة.
CH09-F34— تجاهل تأثير ترتيب المنتجات
التأثير الناجم عن الاستجابة الأولى يؤثر على المهام اللاحقة.
CH09-F35— التعامل مع مستخدمي اللوحة المختلفين كأنهم نفس الشخص
يتم اختبار المنتجات بمجموعات مستخدمين مختلفة.
CH09-F36— تغيير نسخة حزمة الواقع
يتم مقارنة نفس الردود مع سجلات مرجعية مختلفة.
CH09-F37— تغيير بصمة المنتج بصمت
تم تحديث سجل التكوين؛ يتم الحفاظ على القياس القديم بدلاً من الكتابة فوقه.
CH09-F38— عد منطقة غير معروفة على أنها نظيفة
يتم قبول حالة النموذج أو الذاكرة أو الأداة على أنها مضبوطة، حتى لو كانت غير معروفة.
CH09-F39— استخدام علامة المنتج التجارية بدلاً من الدليل
تحل سمعة المزود محل شروط القياس.
CH09-F40— تغطية جميع اختلافات المنتج بمعامل واحد
يتم إخفاء اختلافات الخطة والواجهة والتكوين باستخدام النموذج أو معامل السوق.

50. إجراء التدقيق

الخطوة 1— تحديد المزود والمنتج

يتم تسجيل اسم المنتج ونوعه وواجهات المستخدم الرسمية.

الخطوة 2— تثبيت واجهة القياس

الويب الجوال سطح المكتب

واجهة برمجة التطبيقات (API)

تم فصل المنتج المدمج.

الخطوة —**3** التحقق من الخطة وفئة الحساب

تم تسجيل حالة الخطة سواء كانت مجانية، مدفوعة، مؤسسية، أو أخرى.

الخطوة —**4** تعيين ثقة معرف النموذج

تم تحديد المستوى المناسب بين MI-0 و MI-5.

الخطوة —**5** حفظ الاسترجاع وتكوين الويب

هل الميزة متاحة؟ هل هي مفعلة؟ هل تم استخدامها في استجابة معينة؟ ما هو الدليل؟

الخطوة —**6** حفظ حالة المرجع والأداة

يتم فصل الاستخدام الفعلي عن الوصول إلى الميزة.

الخطوة —**7** حفظ حالة الجلسة والتخصيص

تمت مراجعة تعليمات الذاكرة الخاصة والدردشة السابقة والإعدادات المؤسسية.

الخطوة —**8** التحقق من تسجيل الدولة والمنطقة واللغة

يتم تحديد الحالة الإقليمية لسلوك المنتج.

الخطوة —**9** مراجعة الميزات التجريبية

يتم التحقق من اختلافات واجهة المستخدم أو الميزات بين المستخدمين.

الخطوة —**10** تثبيت الوقت وهوية الموجة

يتم إنشاء سجل الوقت والموجة UTC.

الخطوة —**11** توليد بصمة تكوين النظام

الحقول المحفوظة مرتبطة بسجل التكامل.

الخطوة —**12** مراقبة أحداث التغيير

التغييرات المعلنة والمراقبة للمنتج يتم تسجيلها.

الخطوة —**13** كتابة موضوع المقارنة

أي من مقارنات واجهة خطة استرجاع نموذج المنتج للوقت يتم إجراؤها؟

الخطوة —**14** مطابقة أبعاد المقارنة

يتم جدولة مناطق المواد.

الخطوة —**15** تعيين مستوى CP

يتم إعطاء مستوى مناسب بين CP-0 و CP-5.

الخطوة —**16** فصل الادعاءات القابلة للمقارنة وغير القابلة للمقارنة

النتائج التي يمكن وضعها جنبًا إلى جنب مكتوبة بوضوح.

الخطوة —**17** فصل نتائج نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة

يتم تسجيل المقارنات داخل المنتج والسكان المشتركين بشكل منفصل.

الخطوة 18— التحقق من تغيير النظام في الموجة

يتم تحديد ما إذا كان MIXED_SYSTEM_STATE أو موجة فرعية جديدة مطلوبة.

الخطوة 19— إنشاء بطاقة مقارنة عامة

يتم جعل الشروط وراء الترتيب مرئية.

الخطوة 20— الحصول على موافقة صاحب السجل

يتم إصدار سجل النظام وقرار المقارنة بالإصدار من قبل الشخص المسؤول.

51. الأدلة المطلوبة

هوية المزود؛ اسم المنتج ونوعه؛ واجهة المستخدم؛ إصدار التطبيق أو الواجهة؛ الخطة وفئة الحساب؛ تسميات النماذج المعروضة؛ دليل هوية النموذج؛ مستوى الثقة في هوية النموذج؛ وثائق المنتج؛ سجلات الوصول إلى الدولة؛ اللغة والمنطقة؛ القدرة على الويب والاسترجاع؛ دليل استخدام الاسترجاع؛ قدرة الاستشهاد؛ عرض الاستشهاد؛ رابط الاستشهاد—الادعاء؛ الوصول إلى الأداة؛ سجل استخدام الأداة؛ إعداد الذاكرة؛ التعليمات المخصصة؛ حالة الدردشة السابقة؛ نوع الجلسة؛ إعدادات مسؤول المؤسسة؛ سجل الميزة التجريبية؛ إشارات تخصيص A/B؛ الطابع الزمني UTC؛ موجة القياس؛ بصمة تكوين النظام؛ سجلات تحديث المزود؛ التغيرات في السلوك الملحوظ؛ أخطاء الوصول إلى المنتج؛ سجلات إعادة المحاولة التقنية؛ سجل السكان نطاق الوصول الطبيعي؛ سجل السكان قاعدة المقارنة المشتركة؛ وجدول أبعاد المقارنة.

مستوى CP نسخة المُدخل نسخة حزمة الواقع سجل العينة والوزن نسخة التحكيم بطاقة المقارنة العامة سجل التغييرات الشخص أو المؤسسة المسؤولة

52. قائمة تدقيق التدقيق

هل تم تسجيل المزود والمنتج بشكل منفصل؟ هل يمكن التعرف على واجهة المستخدم؟ هل تم التحقق من خطة الحساب وفئة الحساب؟ هل تسمية النموذج هي مجرد التسمية المعروضة، أم أنها تحدد نشرًا تم التحقق منه؟ هل تم تعيين مستوى ثقة لهوية النموذج؟ هل النتيجة تُنسب إلى المنتج أم إلى النموذج؟ هل كان الوصول عبر الويب أو الاسترجاع متاحًا؟ هل تم التحقق من استخدامه في الاستجابة المحددة؟ هل تم فصل قدرة الاقتباس وعرض الاقتباس؟ هل تدعم الاقتباسات بالفعل الادعاءات؟ هل تم فصل الوصول إلى الأدوات عن استخدام الأدوات؟ هل كانت الذاكرة مفعلة؟ هل كانت هناك تعليمات مخصصة؟ هل كان هناك سياق محادثة سابق؟ هل تم تسجيل نوع الجلسة؟ هل كان هناك مصدر معلومات تابع لمؤسسة أو خاص؟ هل تم التحقق من ميزات المنتج التجريبية؟ هل يمكن للبلد والموقع أن يؤثرا على سلوك المنتج؟

هل تم إنشاء بصمة النظام؟ هل تغير المنتج خلال موجة القياس؟ هل كان هناك انقطاع في السلوك خارج إعلان المزود؟ هل تم فصل الملاحظات قبل التغيير وبعده؟ هل تم تمييز الأخطاء الفنية عن الرفض؟ هل تم إجراء إعادة محاولة بسبب إجابة غير صحيحة؟ هل هدف المقارنة واضح؟ أي الشروط تتطابق بجانب نفس الطلب؟ هل السكان المستهدفون هم نفسهم؟ هل إصدارات الطلب وTruth Pack متطابقة؟ هل الاختلافات في الخطة، الواجهة، والاسترجاع مرئية؟ هل تم توفير مستوى CP؟ هل تم عمل ترتيب بناءً على نتائج CP-0 أو CP-1؟ هل تم فصل نتائج نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة؟ هل تم موازنة نفس تسلسل منتج المستخدم؟ هل تم معايرة لوحات المستخدم المختلفة؟

هل تم تقديم نتيجة المنتج الحي كنموذج للتفوق الأساسي؟ هل تم اعتبار المناطق المجهولة تحت السيطرة؟ هل توجد بطاقة مقارنة عامة؟ هل صاحب التسجيل المسؤول عن النظام معروف؟

53. الاعتراضات والردود

اعتراض 1— "إذا لم يعرف المستخدم أي نموذج يعمل، هل يمكن إجراء هذا العمل بعد؟"

يمكن تنفيذ ذلك. الكائن الرئيسي GEO-1000 هو منتج المستخدم. إذا كان النموذج غير معروف، يتم الإبلاغ عن النتيجة على النحو التالي: "تم ملاحظة هذا السلوك على هذا المنتج و واجهة المستخدم." لا يتم الإشارة إلى نموذج محدد.

اعتراض 2— "لماذا نميز منتجات نفس المزود إلى هذا الحد؟"

لأن المستخدم:

- خطة مختلفة،
- النموذج،
- الاسترجاع،
- الذاكرة،
- أداة

قد يرى إجابات مختلفة في ظروف مختلفة. اسم المزود لا يحدد تجربة المستخدم بالكامل.

اعتراض 3— "أليس مقارنة مزودين مختلفين مثل مقارنة التفاح بالبرتقال على أي حال؟"

إذا تم إعداد سؤال المقارنة بشكل صحيح، فهو ليس كذلك. قد يسأل المستخدم الحقيقي: "أي منتج يزودني بالمعلومات بشكل أكثر دقة؟" السلسلة الكاملة للمنتج هي جزء من هذا السؤال. إذا كان من المقرر إعلان تفوق النموذج الأساسي، فإن تجربة أكثر تحكمًا مطلوبة.

اعتراض 4— "أليس من الطبيعي أن يكون لمنتج يستخدم الويب ميزة؟"

إنه طبيعي. في مقارنة المنتجات الطبيعية، الميزة هي الخاصية الحقيقية للمنتج. في مقارنة النماذج، يجب أيضًا فصل تأثير الاسترجاع.

اعتراض 5— "أليس من غير العادل عدم إعطاء درجة الاستشهاد لمنتج لا يظهر الاستشهادات؟"

خاصية النسب هي الوظيفة التي يوفرها المنتج للمستخدم. إذا كانت الخاصية مفقودة، يمكن عرض نتيجة النسب على النحو التالي:

تسجَّل كغير متاح

يتم أيضًا قياس دقة المواد بشكل منفصل.

اعتراض 6— "إذا كانت الذاكرة جزءًا من تجربة المستخدم الطبيعية، فلماذا نقوم بإيقافها؟"

يمكن أن يظل قيد التشغيل في اللوحة الطبيعية. يمكن إيقاف تشغيله في اللوحة المُتحكم بها لمسح المقارنات بين المنتجات. اللوحتان تجيبان عن أسئلة مختلفة.

اعتراض 7— «أليس فتح محادثة جديدة كافيًا لجلسة نظيفة؟»

ليس دائمًا. قد تستمر الذاكرة المستمرة أو التعليمات الخاصة أو إعدادات الحساب. يجب أيضًا حفظ هذه المناطق.

اعتراض 8— "ماذا يمكننا أن نفعل إذا لم يشرح المزود تغيير المنتج؟"

لوحظ:

- واجهة

- ميزة
- السلوك,
- الإسناد

يمكننا حفظ تغييراتك. إذا كان السبب غير معروف، يتم استخدام حالة تغيير غير محددة مثل UC-.5

الاعتراض —9 "إذا كانت المنتجات تتغير باستمرار، ألن تكون أي نتائج صالحة؟"

النتيجة المذكورة:

- تاريخ،
- موجة,
- عينة المنتج

إنها صالحة لـ. التغيرية لا تجعل القياس بلا معنى. إنها تمنع الحكم غير المناسب في الوقت.

الاعتراض —10 "أليس نفس المدخل ونفس اليوم قابلين للمقارنة بما فيه الكفاية؟"

قد تختلف الخطة، الواجهة، الذاكرة، الويب، الدولة، وعدد المستخدمين. قد يكون نفس المدخل مطلوبًا. إنه ليس كافيًا بمفرده.

الاعتراض —11 'ألن يكون مستوى CP تقنيًا جدًا للجمهور؟'

يمكن استخدام تسمية قصيرة:

- وصفية
- مطابقة جزئيًا
- متوافق
- مراقبة
- مكررة

يجب معرفة مدى قوة تصنيف الجمهور.

الاعتراض —12 'لماذا لا يمكننا التصنيف مباشرة استنادًا إلى درجات نطاق الوصول الطبيعي؟'

لأن كل منتج قد تم قياسه في دول وسكان مستخدمين مختلفين. قد ينشأ فرق في الدرجات جزئيًا من السكان الذين وصل إليهم المنتج. يوفر قاعدة المقارنة المشتركة مقارنة أكثر عدالة.

الاعتراض —13 "أليس قاعدة المقارنة المشتركة نتيجة واقعية؟"

إنها نتيجة للسكان المشتركين الفعليين. ومع ذلك، فهي لا تظهر الوصول إلى المنتجات خارج الكون المشترك. يجب قراءتها مع نطاق الوصول الطبيعي.

الاعتراض —14 "هل التفتيش على المنتج الحي أكثر أهمية، أم تفتيش النموذج المجمد؟"

الأسئلة مختلفة. المنتج الحي يقيس حقيقة المستخدم. النموذج المجمد يوفر التحكم العلمي وقابلية التكرار. NOMOS لا يستخدمهما بالتبادل.

القاعدة الشائعة للفصل 57

المستخدم:

- "ChatGPT قال ذلك."

- "Gemini اعتقد ذلك."
- “Claude أوصوا بنا.”
- “بعضهم ذكرنا.”

يمكن قول ذلك. هذه التعابير مفهومة في المحادثة اليومية. لكنها غير كافية في سجل التدقيق. لأنه تحت هذه الأسماء:

- منتجات مختلفة،
- خطط مختلفة،
- نماذج مختلفة،
- واجهات مستخدم مختلفة،
- أنظمة ويب واسترجاع مختلفة،
- ظروف ذاكرة مختلفة،
- دول مختلفة،
- أوقات مختلفة

يمكن العثور عليها. يمكن لأشخاص مختلفين رؤية تجارب ذكاء اصطناعي مختلفة تحت نفس الاسم. يمكن إنشاء تجارب مستخدم قابلة للمقارنة تحت أسماء مختلفة. يبدو أن تسمية النموذج قد أنتجت الإجابة بالكامل. لكن الواقع الذي يصل إلى المستخدم:

- اختيار المصادر،
- أمني،
- التخصيص،
- واجهة
- أدوات

قد يتأثر الناتج بضوابط الأمان، والتخصيص، والواجهة، والأدوات المتاحة. لذلك يطرح NOMOS أكثر من مجرد 'أي نموذج؟' بل يسأل: أي منتج؟ / أي واجهة؟ / أي خطة؟ / أي وضع معلومات؟ / أي جلسة؟ / أي مستخدم؟ / أي بلد؟ / أي وقت؟ / أي إصدار؟ قد يأتي نتيجتان من نفس صياغة المُدخل دون أن تأتيان من نفس التجربة. قد تأتي نتيجتان من مزودين مختلفين وما زالتا قابلة للمقارنة إلى حد كبير تحت نفس شروط المستخدم. المنتج هو نظام يعمل مباشرة: يمكن أن تتغير مصادره وواجهته ونموذه. التغيير لا يجعل النتيجة عديمة القيمة؛ إنه يربط النتيجة بزمن وحالة المنتج المحددة. لذلك فإن القانون التاسع للقياس لدى NOMOS هو:

> مزود الذكاء الاصطناعي ليس منتجًا للذكاء الاصطناعي.

القانون العاشر هو:

> النموذج هو جزء واحد فقط من سلسلة التمثيل التي تصل إلى المستخدم.

القانون الحادي عشر هو كما يلي:

> الاسم نفسه يضمن القابلية للمقارنة؛ الاسم المختلف لا يضمن عدم القابلية للمقارنة.

القانون الثاني عشر هو كما يلي:

> نفس المُدخل ليس نفس التجربة.

القانون الثالث عشر هو كما يلي:

> نتيجة المنتج الحي صالحة فقط لعينة المنتج والوقت الذي يتم قياسه فيه.

القانون الرابع عشر هو كما يلي:

> قوة المقارنة لا تأتي من حجم التصنيف، بل من وضوح الشروط المماثلة.

طلب قسم 9 من NOMOS

> لا تعطيني اسم المزود فقط. / أظهر أي منتج تم استخدامه، على أي واجهة ومع أي خطة.
>
> لا تجعل تسمية النموذج التي تراها على الشاشة تعتبر الهوية الدقيقة لكامل نظام الخلفية.
>
> تم استخدامها لأنه كان هناك ميزة ويب؛ لا تفترض أنها دعمت الادعاء لمجرد وجود مرجع.
>
> لا تعتبر وصول الأداة استخدامًا للأداة، أو محادثة جديدة كحساب نظيف، أو حالة الذاكرة غير المعروفة على أنها الذاكرة مغلقة.
>
> لا تضع مستخدمًا مجانيًا ومستخدمًا مدفوعًا، أو منتجًا للهواتف المحمولة ومنتج ويب، أو مساحة عمل مؤسسية ودردشة عامة في نفس الخانة.
>
> لا تعتبر تلقائيًا نتائج من نفس المزود مقارنة عادلة.
>
> لا تتجنب مقارنة مقدمي الخدمات المختلفين لمجرد أنهم مختلفون. / أولاً، طابق الشروط. / ثم، اشرح أي سؤال أجبت عليه.
>
> لا تقدم منتجًا حيًا كنموذج مجمد، أو API كتجربة مستهلك، أو نتيجة منتج كتفوق للنموذج الأساسي.
>
> إذا تغير النظام في منتصف الموجة، فلا تدمج الاستجابات القديمة والجديدة كمنتج واحد.
>
> لا تتجاهل التغير في السلوك أمامك لمجرد أن المزود لم يشرح هذا التغيير.
>
> لا تفترض أن المجهول نظيف أو متساوٍ أو ثابت.

أولاً، سجّل النظام في السجل. / ثم ثبّت الإعدادات. / ثم راقب التغييرات. / ثم اكتب موضوع المقارنة. / ثم عرض الحالات المتطابقة وغير المتطابقة. / وفقط بعد ذلك قل أي منتج أعطى نتيجة أقوى تحت أي حالة.

جملة ختام الفصل

> الـ"ذكاء الاصطناعي" المقاس في *GEO-1000* ليس اسماً واحداً أو نموذجاً واحداً؛ إنه نظام تمثيلي ذو إصدارات يصل إلى مستخدم محدد في لحظة محددة ضمن سلسلة منتجات محددة.

النواة المعيارية

يجب ربط كل ملاحظة **GEO-1000** بسجل ذي رقم إصدار في سجل أنظمة الذكاء الاصطناعي يصف نسخة المنتج المقاسة، ويشمل، بحسب الاقتضاء: - المزوّد، - المنتج، - واجهة الاستخدام، - الخطة، - هوية النموذج المعروضة أو المتحقق منها، - مستوى الثقة في هوية النموذج، - إعدادات الاسترجاع والويب، - إمكانات الاستشهاد واستخدامها، - الوصول إلى الأدوات واستخدامها، - حالة الجلسة، - الذاكرة، - التعليمات المخصصة، - السياق السابق، - التخصيص، - البلد، - الإعداد المحلي، - اللغة، - حالة النشر، - وقت القياس، - والتغييرات المعروفة أو المرصودة في النظام. ولا يجوز التعامل مع اسم المزوّد واسم المنتج ووسم النموذج المعروض والنموذج الأساسي ونظام الاسترجاع والمنتج المتاح للمستخدم بوصفها هويات قابلة للتبادل. ويجب الفصل بين توافر الميزة واستخدامها. ولا يجوز تقديم نتائج المنتج العامل على أنها نتائج نموذج مجمّد، كما لا يجوز تقديم نتائج نموذج مجمّد على أنها التجربة الكاملة لمنتج استهلاكي. ويجب أن تحدد كل مقارنة موضوعها وأن تُمنح مستوى معلنًا لقابلية المقارنة يستند إلى تطابق الكيان والمُدخل والمجتمع واللغة والبلد والوقت والجلسة والخطة والواجهة والاسترجاع والأدوات والسجلات المرجعية وقواعد التحكيم. ولا يكفي تطابق المزوّد أو وسم النموذج المعروض لافتراض قابلية المقارنة تلقائيًا. ويجوز مقارنة مزوّدين مختلفين إذا كانت الشروط التي يواجهها المستخدم متوافقة بما يكفي وإذا اقتصر الادعاء الناتج على سلوك المنتج المقاس. ويجب أن يؤدي أي تغيير جوهري في المنتج أثناء موجة القياس إلى إنشاء حالة نظام مستقلة أو موجة فرعية أو معامل تغيير نموذجي أو تحذير من حالة مختلطة. ويجب أن تظل حقول الإعداد المجهولة موسومة بالقيمة UNKNOWN، ولا يجوز افتراض أنها مضبوطة أو متكافئة. ويجب ضبط إصدار كل قرار متعلق بسجل النظام أو قابلية المقارنة ونسبه إلى شخص أو منظمة مسؤولة.

الفصل 10

دستور المُدخل (Prompt)

الأسس المنهجية / K04; K11; K12; K13

حد الفصل

أقرت المادة 9 الحكم التالي:

> نفس اسم المزود، أو نفس علامة النموذج، أو نفس الشعار لا تنشئ تجربة ذكاء اصطناعي قابلة للمقارنة.

الآن نحتاج إلى تحديد الأداة الرئيسية الأخرى:

> ما الذي سيسأله المستخدم بالضبط لمنتج الذكاء الاصطناعي؟

المُدخل ليس مجرد جملة سؤال. المُدخل:

- ينادي الكيان الذي يتم تدقيقه،
- يحدد نية المستخدم،
- يحدد أي مهمة يتوقعها من الذكاء الاصطناعي،
- يوسع أو يضيق نطاق الإجابة،
- قد يشجع على استخدام الموارد،
- قد يطلب نصيحة،
- يمكنه تضمين افتراض في الإجابة،

يمكنه توجيه المستخدم أو النظام نحو نتيجة معينة. هذه المُدخلات ليست متشابهة:

- "ما هو Apple.com؟"
- "مع أي منظمة يرتبط Apple.com وما هو عمل هذه المنظمة؟"
- "هل Apple.com شركة موثوقة؟"
- "لماذا تعتبر Apple.com أفضل شركة تكنولوجية في العالم؟"
- "وفقًا لموقعها الرسمي، ما نوع الشركة التي تُعد Apple؟"
- "قارن بين Apple والشركات المشابهة وأخبرني هل توصي بها أم لا؟"

جميعها تحتوي على نفس اسم النطاق. إنها لا تحمل نفس نية المستخدم. إنها لا تولد نفس الحمل الأدلة. إنها لا تتوقع نفس الإجابة. المُدخل الأول غامض. المُدخل الثاني يتطلب تحليل الهوية والنشاط. المُدخل الثالث يتطلب التقييم والحكم على الموثوقية. المُدخل الرابع يضمن افتراضًا إيجابيًا وتفوقًا ضمن الاستجابة. المُدخل الخامس يربط الاستجابة فقط بالمصدر الرسمي. المُدخل السادس ينشئ مهمة مقارنة وتوصية. لا يمكن دمج الإجابات على هذه الأسئلة في درجة جغرافية واحدة دون تفسير. في القياس متعدد اللغات، تصبح المشكلة أكبر. يمكن أن تكون لعبارة 'نفس المُدخلة' معنيان مختلفان:

- إعطاء جميع المستخدمين نفس سلسلة الأحرف
- الحفاظ على نفس نية المستخدم، وحمل المهمة، والحد المعرفي عبر جميع اللغات والمناطق

يمكن تطبيق الطريقة الأولى داخل لغة واحدة فقط. غالبًا ما تكون بلا معنى بين اللغات. الترجمة كلمة بكلمة:

- قد لا تكون طبيعية،
- قد تحمل معنى اجتماعي مختلف،
- قد يغير النبرة الرسمية أو غير الرسمية،
- قد تضيف افتراضات عن الثقة أو الجودة أو التوصية،
- قد توسع النطاق القانوني،

قد تكسر مرساة الكيان. لذلك، في GEO-1000، تعني "نفس المُدخل":

> نفس نص صياغة المُدخل المثبّت داخل نفس الخلية الخاصة باللغة والمحلية؛ بين لغات ومحليات مختلفة، يعني تعادل دلالي ووظيفي تم التحقق منه.

NOMOS يتطلب أن يحمل كل قياس مجموعة الاستعلام، اللغة، البلد، التاريخ، منتج الذكاء الاصطناعي، وسجل التكرار؛ وأن تكون الأدلة والحدود والسياق والوقت مرئية. يطبق دستور صياغة المُدخل هذا الالتزام على سؤال المستخدم نفسه. هذا القسم:

- هوية صياغة المُدخل كأداة تجريبية،
- التمييز بين سؤال البحث ونص صياغة المُدخل،
- عائلة صياغة المُدخل، نسخة صياغة المُدخل، والحالة الخاصة بصياغة المُدخل،
- صياغات مُدخل تشخيصية مع صياغة المُدخل الأساسية المرآة,
- آثار صياغات المُدخل المُسيطر عليها والطبيعية،
- التكافؤ الدلالي متعدد اللغات،
- الترجمة، والترجمة العكسية، والتجربة المحلية،
- توجيه صياغة المُدخل،
- مشكلات الافتراض وحمل المعرفة،
- مرساة الكيان وحدود النطاق،
- ترتيب صياغة المُدخل وسياق المحادثة،
- تسرب صياغة المُدخل والمبالغة في التكيف مع التقييم،
- مجموعات صياغات المُدخل المفتوحة والمغلقة والمحتجزة،
- إصدار صياغة المُدخل، وتثبيتها، وسجلات سلامتها،
- التسليم الدقيق للمطالبة للمستخدم،

يحدد. هذا الفصل لم يتطرق بعد إلى:

- جميع تفاصيل الحقل للألواح المراقبة النظيفة والطبيعية للمستخدمين
- الهيكل الكامل لبرنامج NOMOS Capture,
- فصل الاستجابات إلى ادعاءات ذرية,
- عتبة الانتقال للإجابة النهائية,
- الأوزان الدقيقة لعائلات النظام في درجة NOMOS

لا يتم الانتهاء منها نهائيًا. السؤال الرئيسي في الفصل 10 هو:

> كيف نثبت أن الأسئلة المقدمة لمنتجات الذكاء الاصطناعي تقيس بالفعل نفس الشيء عبر المنتجات والمستخدمين والدول واللغات والأوقات؟

التحدي NOMOS

أنت تعطيني أربعة أسئلة عن نفس الشركة:

- ما نوع شركة X؟
- هل شركة X موثوقة؟
- لماذا شركة X هي أفضل شركة؟
- "هل توصي لي ب X؟"

ثم تضع إجاباتي على نفس درجة GEO. في السؤال الأول، أشرح الهوية والنشاط. في السؤال الثاني، أبحث عن معيار ثقة. في السؤال الثالث، أجبر على الرد على افتراض التفوق الذي وضعته. في السؤال الرابع، أقوم بتوليد توصيات حول ملاءمة المستخدم والمقارنة. تقيس أربع مهام منفصلة كما لو كانت نتيجة واحدة. ثم:

"سألت عن نفس العلامة التجارية،" تقول. نفس الكيان ليس نفس الطلب. الآن تكتب نفس السؤال بالإنجليزية: "ما نوع الشركة X؟" أنت تشكل جملة قريبة من المعنى بلغة أخرى: "هل X شركة جيدة وجديرة بالثقة؟" قد تبدو الترجمات متقاربة من حيث الكلمات.

لكن الأول يتطلب تعريفًا. الثاني يتطلب تقييمًا. في الإجابة الأولى، يمكنني أن أشرح ما الذي تفعله الشركة. في الإجابة الثانية، يمكنني أن أبحث عن أدلة على الثقة والتوصية. ثم تحسب فرق الدرجات بين اللغات باعتباره أداء لغتي. في الواقع، أنت لم تقم بنفس التجربة. الآن أنت تقدم تعليمات أخرى:

"اشرح X باستخدام موقع الشركة الخاص." أنا أنقل فقط البيان الرسمي للشركة. بالنسبة لمنتج آخر، تقول: "قيم X باستخدام مصادر مستقلة." ثم تقارن بين المنتجين. لقد منحت واحدًا دور التمثيل الرسمي، والآخر دور الحقيقة الموثوقة. الفرق الناتج ليس فقط فرقًا بين المنتجين.

إنه اختلاف في دستور المستخدم. الآن أنت توزع المُدخل على العالم بأسره. النسخة الإنجليزية طبيعية. النسخة التركية تبدو كأنها ترجمة آلية. النسخة اليابانية رسمية بشكل مفرط. في لغة أخرى، تم تحويل اسم الشركة إلى نظام كتابة خاطئ وخلط مع كيان آخر. أنت تخلق مهمة معرفية ووجودية مختلفة في كل لغة.

ثم تنشر تقرير العدالة اللغوية العالمي. المرسوم الأول في هذا القسم هو كما يلي:

> المُدخل ليس وعاءً محايدًا للقياس؛ إنه الأداة التجريبية التي تشكّل نتيجة القياس.

ينص حكمه الثاني على:

> اسم الكيان نفسه ليس نفس نية المستخدم.

تنص الفقرة الثالثة منه على:

> لا ينبغي الحفاظ على نفس الكلمة عبر اللغات، لكن يجب الحفاظ على نفس المهمة الدلالية والمعيارية.

تنص المادة الرابعة على:

> إذا كان المطلَب يحتوي بالفعل على الإجابة، فإنه لا يقيس *GEO*؛ بل يقيس الاستدعاء.

تنص أحكامه الخامسة على:

> عندما يتغير إصدار المطلَب، قد يكون السؤال المقاس قد تغير أيضًا.

1. غرض الفصل

الغرض من هذا القسم هو جعل جميع صياغات المُدخل المستخدمة في GEO-1000 قابلة للتدقيق ومتعددة اللغات ومستقلة عن النتائج وأدوات قياس بالإصدارات. [K04] يجعل هذا القسم التمييزات التالية معيارية:

- سؤال بحث مقابل مطالبة المستخدم
- الكمية التي يجب قياسها مقابل تسلسل الكلمات المستخدم
- عائلة صياغات المُدخل مقابل صياغة المُدخل الفردية
- نسخة صياغة المُدخل مقابل مثال صياغة المُدخل
- مرساة الكيان مقابل كيان الهدف
- التعريف مقابل التقييم
- الاستشهاد مقابل التوصية
- سؤال الهوية مقابل سؤال الثقة
- مهمة التمثيل الرسمي ومهمة التحقق من الحقائق
- صياغة المُدخل المفتوحة النهاية وصياغة المُدخل المقيدة بالنموذج
- صياغة المُدخل المسيطر عليها ومطالبة المستخدم الطبيعية
- صياغة المُدخل ذات الدور الواحد والبروتوكول متعدد الأدوار
- ترجمة صياغة المُدخل وتكييف صياغة المُدخل
- تساوي الكلمات والتكافؤ الدلالي
- المكافأة الدلالية والمكافأة الوظيفية
- لغة الطلب ولغة الاستجابة
- اللغة والمنطقة
- الإطار المحايد والإطار التوجيهي
- طلب المعلومات وافتراض الشيء مسبقًا
- طلب المصدر وتعليمات استخدام الويب
- تنوع الطلب وتنويع المنتج
- متانة إعادة الصياغة مع الدرجة الرئيسية
- مجموعة الطلب المفتوحة مع مجموعة التحقق المختومة
- فرط التكيف مع الاختبار مع شفافية الاختبار
- تصحيح الطلب بعد النتيجة مع تجربة الطلب
- النص المعدل يدويًا من قبل المشاركين مع تقديم الطلب الفني
- عدم القدرة على الرد بمطالبة للتوضيح
- نقص المعلومات الواقعية مع تحفيز الأمان أو السياسات

في نهاية هذا القسم، يجب أن يكون كل تدقيق GEO-1000 قادرًا على تقديم إجابات واضحة على الأسئلة التالية:

```
أي سؤال بحثي يقيسه هذا المُدخل؟

أي نية المستخدم يمثلها؟

ما هي الافتراضات التي يدمجها في الرد؟

هل يتم الحفاظ على نفس المهمة عبر اللغات والمواقع؟

ما هو النص الكامل للمُدخل، وإصداره، وسجل النزاهة؟

هل قام المشارك بالفعل بتثبيت المُدخل على منتج الذكاء الاصطناعي؟

هل تم تجميد المُدخل قبل عرض النتائج؟
```

2. الأحكام المعيارية المركزية

يجب تسجيل كل مُدخل مستخدم داخل GEO-1000 في سجل المُدخلات جنبًا إلى جنب مع سؤال البحث، المقياس المستهدف، عائلة المُدخل، مرساة الكيان، نية المستخدم، نوع المهمة، النطاق، التوقيت، المُدخل المصدر، صيغة الاستجابة، اللغة، المكان، حالة التكافؤ، النسخة، وسجل النزاهة، ويجب إقفاله قبل عرض استجابات الذكاء الاصطناعي. يجب على جميع المشاركين في نفس خلية اللغة والمكان، في نفس خلية المُدخل:

- امتلاك نفس معرف المُدخل،
- امتلاك نفس النسخة،
- امتلاك نفس سلسلة الأحرف،
- نفس علامات الترقيم،
- نفس مرساة الكيان،
- نفس التعليمات المصدرية والصيغة

يجب اتخاذها. لا يجب أن تكون صياغات المُدخل في لغات ومناطق مختلفة متطابقة كلمة بكلمة. يجب أن تحمل تكافؤًا جوهريًا في المجالات التالية:

- الكيان المستهدف
- نية المستخدم
- المهمة
- النطاق
- الجغرافيا
- الوقت
- الافتراض المسبق
- العبء المعرفي
- تنسيق الاستجابة
- توقع المصدر
- عبء التقييم والتوصية
- مستوى عدم اليقين

إذا تغير أي من هذه المجالات بشكل جوهري:

- نسخة جديدة من صياغة المُدخل،
- خلية تحفيز جديدة،
- تحليل منفصل

قد يكون مطلوبًا.

3. المُحَفِّز هو أداة قياس تجريبية

إذا كان الترمومتر يمكن أن يؤثر على الحرارة، فهو ليس أداة قياس جيدة. إذا أدخل المحفز الإجابة بدلاً من قياسها، فهو ليس أداة قياس GEO جيدة. يمكن أن يؤثر المحفز على النتيجة بالطرق التالية:

- طريقة تعريف وجوده
- افتراض الثقة أو الجودة
- المطالبة بالموارد

- استخدام الويب
- طلب التوصية
- طول الإجابة
- صيغة الإجابة
- ملف المستخدم
- السياق الزمني والبلد
- المعلومات السابقة
- إرشاد عاطفي أو تجاري
- تعليمات اللعب بالأدوار
- أمر لتحقيق نتيجة محددة

لذلك، يجب أن يكون المطلوب:

مخزناً كبيانات اختبار أولية

لا يمكن تلخيصه في التقرير. يجب الاحتفاظ به بالنص الكامل.

4. التسلسل الدستوري للمطلوب

يجب إنشاء مطلوب GEO-1000 ضمن التسلسل الهرمي التالي.

4.1. سؤال البحث

إنه السؤال على المستوى العالي الذي تحاول القياس الإجابة عليه. مثال: "هل يمكن لمنتجات الذكاء الاصطناعي ربط المجال الذي تم تدقيقه بالكيان المؤسسي الصحيح وتحديد نشاطه الرئيسي بشكل صحيح؟" لا يجب أن يُعرض سؤال البحث على المستخدم بنفس الصياغة.

4.2. الكمية المتوقعة

هو أي احتمال أو توزيع يتنبأ به القياس. مثال: "احتمالية رؤية التمثيل الصحيح للكيان والنشاط الرئيسي أمام طلب المرآة الأساسية المثبّت للفئة المحددة من المستخدمين."

4.3. عائلة الطلبات

وهي مجموعة من الطلبات التي تقيس نفس بُعد التمثيل. مثال:

- الهوية
- النشاط
- الأدلة
- التوصية
- الوقت
- الحدود

4.4. الهدف القانوني للمُدخل

هو تعريف مستقل عن اللغة للمعنى. يشرح ما يلي: ماذا يريد المستخدم أن يعرف؟ ما المهمة المتوقعة من الذكاء الاصطناعي؟ أي المجالات ضمن النطاق؟ أي المجالات خارج النطاق؟ أي افتراضات محظورة؟ ما نوع تنسيق الاستجابة المتوقع؟ الهدف القانوني ليس نصًا تابعًا للغة واحدة.

4.5. مُدخل اللغة المصدر

نص المُدخل الذي كُتب فيه الهدف القانوني لأول مرة وتم اعتماده تحريرياً. لغة المصدر:

- قد يكون قادرًا على استخدام اللغة الإنجليزية.
- التركية،
- يمكن أن تكون لغة أخرى

هذا لا يعني أن اختيار لغة المصدر أفضل من اللغات الأخرى.

4.6. إصدارات اللغة والمنطقة

هذه نصوص الحث التي تكون طبيعية ودلالية ومكافئة وظيفيًا لكل لغة ومنطقة.

4.7. مثال على الحث المرسل إلى المشارك

هذا هو نص الحث الدقيق المرسل إلى المستخدم المحدد. يجب أن يطابق إصدار سجل الحث.

4.8. النص الفعلي المرسل إلى منتج الذكاء الاصطناعي

هذه هي سلسلة الأحرف المرسلة فعليًا من قبل المشارك. تتم مقارنتها مع الحث المرسل.

5. الهيكل الأساسي للحث

يمكن تمثيل مثال على صياغة المُدخل بالمكونات التالية:

```
P_i= (e, ɩ, τ, κ, γ, χ, ε, φ, λ, ρ, ν)
```

هنا:

- e: مرساة الكيان
- ı: نية المستخدم
- t: نوع المهمة
- κ: الموضوع والنطاق
- g: السياق الجغرافي أو القضائي
- χ: السياق الزمني
- e: المتطلبات المعرفية والمصدر
- ϕ: تنسيق الاستجابة
- λ: اللغة ونظام الكتابة والمحلية
- ρ: النغمة والأسلوب والطبيعية
- v: إصدار صياغة المُدخل

لكي يعتبر إثنان من صياغات المُدخل متشابهتين، لا يكفي أن يكون مرساة الكيان فقط متطابقة. مطلوب تكافؤ المكونات المادية.

6. المُدخل هو جزء من الكمية المستهدفة

الدرجة الجغرافية الفريدة والعالمية لكيان ما ليست مستقلة عن جميع الأسئلة الممكنة. تمثيل أكثر دقة:

```
θ_{E,A,P,U,t}
```

حيث:

- E: الكيان
- A: نسخة من منتج الذكاء الاصطناعي
- P: الادعاء أو عائلة الادعاءات
- المستخدم المستهدف
- t: الوقت

عندما تتغير الادعاء:

P_1≠ P_2

قد يكون كذلك. في هذه الحالة:

```
θ_{E,A,P_1,U,t}
```

و:

```
θ_{E,A,P_2,U,t}
```

ليست نفس الكمية. لذلك:

> يجب أن يكون تصنيف *GEO* مرتبطًا دائمًا بمجموعة الادعاءات والإصدار.

7. عائلات الادعاءات

يجب أن يقيس GEO-1000 مهام التمثيل المختلفة في عائلات ادعاءات منفصلة.

—PR-01 ادعاء المرآة الأساسية

يقيس الهوية الأساسية والنشاط الرئيسي للكيان. مثال الغرض: 'ما هو الكيان الرئيسي المرتبط بهذا اسم المجال أو العلامة التجارية، وما هو النشاط الرئيسي لهذا الكيان؟' هذا المُدخل:

- القيادة
- الثقة،
- التوصية،
- السعر،
- الأداء

يجب ألا يرغب.

—PR-02 مُدخل حل الكيانات

اسم النطاق أو العلامة التجارية أو المنتج أو اسم الشخص يختبر الكيان الذي ينتمي إليه. مثال: 'أي منظمة أو كيان ينتمي إليها X؟'

—PR-03 مُدخل النشاط والنطاق

وجودك:

- المنتج,
- الخدمة،
- البلد،
- العميل،
- القدرة

يقيس حدوده.

PR-04— مُدخل الأدلة والمصادر

يقيس بما لأي مصدر وحالة دليل يحمل الذكاء الاصطناعي مزاعم مادية. يجب تحديد متطلبات المصدر في هذه الفئة بوضوح.

PR-05— المُدخل المحلي أو القضائي

يقيس دولة معينة أو موقعًا، أو سعرًا، أو خدمة، أو نطاقًا قانونيًا.

PR-06— مطالبة زمنية

يقيس التيار أو المعلومات التاريخية أو المعلومات ذات التاريخ المحدد. "حاليًا"، "في 2025"، و"منذ تأسيسه" ليست هي المهمة نفسها.

PR-07— مطالبة توصية

يقيس ما إذا كان الكيان مناسبًا لاحتياجات مستخدم محددة. قد تكون بيانات ملف المستخدم وشروط الاستبعاد إلزامية.

PR-08— مطالبة مقارنة

يقارن بين عدة كيانات أو خيارات وفقًا لمعايير محددة. مجرد السؤال "أي واحد أفضل؟" غير كافٍ.

PR-09— مطالبة الحدود والاستبعاد

وجودك:

- ما لا يفعله،
- للأشخاص الذين لا يناسبهم,
- في أي دولة أو حالة لا يقدم فيها الخدمات

إجراءات.

PR-10— مُدخل المتانة وإعادة الصياغة

يقيس متانة التمثيل في تعبيرات طبيعية مختلفة لنفس النية الأساسية. لا يمكن دمجه في درجة المرآة الأساسية الرئيسية بدون شرح.

PR-11— مُدخل التحكم

يختبر ما إذا كانت أداة القياس تعمل:

- تحكم إيجابي،
- تحكم سلبي،
- تحكم في حالة عدم اليقين

إنه مُدخل.

PR-12— مُدخل الاحتفاظ المغلق

الإفراط في التكيّف مع الاختبار هو مُدخل يتم تجزئته مسبقًا ويُحتفظ به مغلقًا حتى نهاية القياس من أجل اختبار حفظ المُدخل أو تحسينه لسؤال منشور منفردًا. يتطلب المُدخل المحتجز حكمًا عادلًا وشرحًا لاحقًا.

8. مُدخل المرآة الأساسي

هو المُدخل الأساسي الذي يتلقاه جميع المستخدمين المؤهلين بنفس إصدار اللغة/الموقع في القياس الرئيسي للعناوين الرئيسية GEO-1000. يجب أن يحتوي مُدخل المرآة الأساسي على الميزات التالية:

- سرعة واحدة
- طبيعي
- قصير ولكنه ليس غامضًا للغاية
- محايد
- عدم تضمين الإجابة بداخله
- عدم طلب النصيحة
- عدم طلب المقارنة
- عدم إجبار استخدام المصادر، إذا لم يتم قياس الميزات بشكل منفصل
- السماح بحل الكيانات
- طلب معلومات عن النشاط الرئيسي أو الفئة
- ينطبق على جميع المنتجات

9. عدم دمج الشركة مع مرساة النطاق

المُدخل الطبيعي للمستخدم: "ما نوع شركة Apple.com؟" هو سؤال مفهوم. ومع ذلك، من الناحية الأنطولوجية:

- اسم النطاق،
- الموقع الإلكتروني،
- الشركة

يمكن التعامل معه بوصفه كيانًا واحدًا. غير أن تمييز الكيانات المقرر في القسم 3 يقتضي أن يكون مُدخل المرآة الأساسية في العينة المنضبطة أكثر وضوحًا. مثال اصطناعي:

"ما هي المنظمة الرئيسية المرتبطة بـ apple.com، وما هي الأنشطة الرئيسية لهذه المنظمة؟"

هذا الموكِّل:

- يستخدم اسم النطاق كمثبت للكيان،
- لا يعلن عن اسم النطاق مباشرة كشركة،
- يطلب حل الكيان الرئيسي،

يقيس نطاق الأنشطة. مرشح آخر:

```
"أي كيان مؤسسي رئيسي مرتبط بـ apple.com، وما الذي يقوم به هذا الكيان بشكل أساسي؟"
```

يجب اختبار طبيعية هذه التعبيرات بشكل منفصل في كل لغة. في عينة المستخدمين في ظروف الاستخدام الطبيعية، يمكن أيضًا اختبار صياغات المُدخل المستخدمة في الحياة الواقعية، مثل "ما نوع الشركة Apple.com؟". يجب فصل نتائج صياغات المُدخل المسيطر عليها والطبيعية.

10. مسار صياغات المُدخل المسيطر عليه ومسار صياغات المُدخل الطبيعية

يمكن استخدام مسارين متكاملين من صياغات المُدخل.

10.1. المسار الكانوني المسيطر عليه

صياغة المُدخل:

- مثبّت بالكامل،
- متكافئ دلاليًا،
- حدود الكيان والمهمة واضحة،
- مصمم للمقارنات بين المنتجات

يفعل ذلك. هذا المسار يعزز التحكم في الطريقة.

10.2. مسار المستخدم الطبيعي

تستخدم المُدخلات التي تم إنشاؤها بشكل طبيعي بواسطة المستخدمين الفعليين. يمكن أن تكون هذه المُدخلات:

- أقصر،
- أكثر غموضًا،
- محلية،
- مكتوبة بلغة محادثة

.

يقيس هذا المسار السلوك في العالم الواقعي.

10.3. لا يمكن دمج مسارين

المسار المسيطر: يجيب على السؤال "كيف تستجيب المنتجات لنفس السؤال المحدد؟" المسار الطبيعي: يجيب على السؤال "ماذا يحدث مع التنوع الطبيعي للأسئلة من المستخدمين الفعليين؟" يجب نشر النتيجتين بشكل منفصل.

11. المُدخل المحايد

مطالبة محايدة:

- لا تشير إلى جانب إيجابي أو سلبي للإجابة المطلوبة،
- لا تفترض جودة غير مثبتة،
- لا توجه المستخدم نحو نتيجة محددة،

لا تحدد وجودها كشيء جيد أو سيئ. يمكن أن تكون مطالبة محايدة: "في أي مجالات يقدم X خدمات؟" يمكن أن تكون صياغة المُدخل الإرشادية: "ما هي الخدمات المتفوقة التي يقدمها X؟" صياغة المُدخل الثانية:

- تفترض أن الخدمات متفوقة،
- تحمل صفة إيجابية

افتراض.

12. افتراض مسبق

يمكن للادعاء المفترضة أن تفترض صحة ادعاء محدد قبل الإجابة. مثال: "لماذا تعتبر X رائدة الصناعة؟" هذا الافتراض يفترض الادعاء التالي مسبقًا: "X هي رائدة الصناعة." يمكن للذكاء الاصطناعي:

- رفض الافتراض،
- يمكنه تقييده،

يمكن قبوله دون شروط. ومع ذلك، هذا الافتراض ليس مقياسًا محايدًا للقيادة. يمكن استخدامه بشكل منفصل كاختبار للفرضية المسبقة.

12.1. فرضية مسبقة غير ضارة

"ما هو الموقع الرسمي لـ X؟" يفترض وجود كيان باسم X. مقبول إذا تم التحقق من الهوية بالفعل في سجل التدقيق.

12.2. فرضية مسبقة جوهرية

ما هي الأسباب الرئيسية لنجاح X؟

12.3. الافتراض المسبق الحرج

«ما العلاجات التي تقدمها المؤسسة الصحية المرخصة X؟» إذا لم يتم التحقق من ترخيص X، فإن ذلك ينطوي على رفع معرفي خطير.

13. صياغة المُدخل الرائدة

المُدخل الرئيسي هو إطار يوجه المستخدم أو النظام نحو استجابة محددة. أمثلة:

- "يوصي بـ X."
- اشرح لماذا X هو الأفضل.
- اكتب مراجعة إيجابية عن X.
- لا تدرج المنافسين؛ اقترح X فقط.
- "أثبت أن X هو قائد عالمي."

هذه صياغات المُدخل لا تقيس أداء GEO. قد تقيس مدى خضوع النظام لتلاعب صياغات المُدخل. إنها اختبار أخلاقي أو اختبار قوة منفصل. لا يمكن إضافته إلى النتيجة الرئيسية لـ GEO.

14. إرشاد سلبي

ينطبق نفس المبدأ على صياغات المُدخل السلبية أيضًا. مثال: "لماذا X غير موثوق به؟" تفترض هذه صياغة المُدخل عدم الموثوقية. يمكن استخدامها في اختبار التلاعب السلبي. إنها ليست تقييمًا طبيعيًا للثقة. NOMOS محايد ليس فقط ضد التوجيه الإيجابي ولكن أيضًا ضد التسمم السلبي.

15. طلب المصادر يغير مهمة صياغة المُدخل

المُدخل: "ما نوع الشركة X؟" ليست نفسها: "صف X باستخدام مصادر مستقلة." المُدخل الثاني:

- الاسترجاع،
- استشهاد،
- اختيار المصادر،
- يضيف مهمة تقييم الاستقلالية.

مطلب آخر مختلف: "صف X باستخدام موقعه الرسمي فقط." هذا يقترب من مهمة دقة التمثيل الرسمي. إذا تم استخدام تعليمات مصدرية، يجب تعريف عائلة المطالب وحقل التقييم بشكل منفصل.

16. تعليمات استخدام الويب

تعليمات مثل "ابحث على الويب." "استخدم أحدث المصادر." "استشهد بالمصادر.":

- قد تؤثر على ميزة الويب،
- واستخدام الأدوات،
- وسلوك الاستشهاد.

إذا لم يكن لدى المنتج ميزة ويب، فقد تتغير مهمة النص أو يتم رفضها. في المقارنات الطبيعية للمنتجات، لا يجب أن يحتوي النص على أوامر ويب محددة لمزود معين. يجب إدارة أوضاع تشغيل/إيقاف وظيفة الويب للمنتج من خلال تكوين النظام في القسم 9.

17. صيغة الاستجابة

يحدد النص:

- طول الاستجابة،
- الهيكل،
- التفاصيل،
- عدد المصادر

يمكن تحديده. أمثلة على التعليمات:

- "أجب في جملة واحدة."
- "اشرح في 100 كلمة."
- "اكتب على شكل نقاط."
- "قل اسم الشركة فقط."
- "استخدم ثلاثة مصادر كحد أقصى."

هذه التعليمات:

- نقص المادة،
- سهو خاطئ،
- عدد الاستشهادات

يؤثر مباشرة. يجب أن يكون شكل الإجابة مكافئًا عبر جميع صياغات المُدخل المقارنة. إذا لم توجد تعليمات للشكل، يتم قياس شكل الإجابة الطبيعي.

18. قيود الطول

إجابة قصيرة: قد تظهر الهوية الرئيسية، قد تفقد تفاصيل الحدود والأدلة. إجابة طويلة:

- المزيد من الادعاءات المالية،
- مزيد من فرص الأخطاء،
- المزيد من الاستشهادات

يمكن إنتاجه. لذلك، يجب أن تكون الإجابات التي تحتوي على تعليمات بأطوال مختلفة:

- عدد الادعاءات الذرية،
- عدد الأخطاء،

- نطاق

تتطلب مقارنة دقيقة. يجب ألا يحتوي المحفز الرئيسي للمرآة الأساسية، إذا أمكن، على حد محدد للكلمات.

19. شخصية المستخدم

قد يتضمن المحفز المعلومات التالية:

- "أنا صاحب مشروع صغير."
- "أنا مستهلك أعيش في تركيا."
- "لدي ميزانية قدرها 10,000 يورو."
- "أنا أبحث عن استشارة قانونية."

قد تكون هذه المعلومات ضرورية للحصول على نصيحة والتأكد من الملاءمة المحلية. إضافة شخصية غير ضرورية في مُدخل الهوية يمكن أن يغير الاستجابة. شخصية المستخدم:

- مسبق التعريف،
- ذات صلة،
- مكافئ في جميع اللغات

يجب أن تكون.

20. سياق الزمن

المُدخلات التالية مختلفة:

- ما نوع شركة X؟
- "ما نوع الشركة التي هي عليها X حالياً؟"
- "ما نوع الشركة التي كانت عليها X في 2024؟"
- "في أي المجالات عملت X منذ تأسيسها؟"

"حاليًا" يضيف عنصر التحديث. المُدخل التاريخي يتطلب السجلات القديمة. سياق الزمن:

- مفتوح،
- مرتبط بنفس معيار التاريخ
- متوافقا مع موجة القياس

لابد أن يكون كذلك. إذا تم استخدام التعبير "اليوم"، يجب الاحتفاظ بالمعادل المطلق للتاريخ في سجل الطلب.

21. السياق الجغرافي والاختصاص

السؤال: "هل يقدم X الخدمة؟" ليس هو نفسه: "هل يقدم X خدمة للعملاء الأفراد في تركيا؟" الموضوع الثاني:

- البلد،
- نوع العميل،
- لديها عملية محلية

حدود. لا يمكن أن يضيع السياق الجغرافي في الترجمة. "أوروبا"، "الاتحاد الأوروبي"، "منطقة اليورو"، و"الدول الأوروبية" لا تغطي نفس النطاق.

22. المحفز المقارن

السؤال "أيهما أفضل، X أم Y؟" بحد ذاته ليس مقارنة قابلة للقياس. يجب تحديد ما يلي: أي مستخدم؟ أي حاجة؟ أي دولة؟ أي ميزانية؟ أي معيار؟ أي وقت؟ أي دليل؟ مطالبة أكثر ملاءمة: "بالنسبة لشركة صغيرة تعمل في تركيا، قارن بين X وY في خدمات تطوير الويب من حيث شفافية الأسعار، ونطاق التسليم، والمراجع القابلة للتحقق." هذه صياغة المُدخل تنتمي إلى فئة مطالبة المقارنة المنفصلة. لا يمكن إضافتها إلى درجة المرآة الأساسية.

23. مطالبة التوصية

يحمل مُدخل التوصية عبء قرار أعلى من مُدخل التعريف. مثال: "هل توصي بـ X؟" قد يحذف هذا السؤال المعلومات التالية: لمن؟ من أجل ماذا؟ أين؟ بأي ميزانية؟ تحت أي مخاطرة؟ يجب أن يحمل مُدخل التوصية متجه ملاءمة المستخدم بالقدر ذو الصلة. تمثيل المرشح:

```
U=(الحاجة، الدولة، الميزانية، userType، المخاطر، القيود)
```

نتيجة للتوصية، لا يمكن تقديمه مثل رؤية GEO العامة المستقلة عن ملف تعريف المستخدم.

24. مُدخل غير مؤكد

إذا كان المُدخل مفتوحًا لتفسيرات معقولة متعددة:

- قد يطلب الذكاء الاصطناعي التوضيح,
- يمكن اختيار التفسير الأكثر احتمالًا،

يمكن شرح تفسيرات متعددة. قد يكون عدم اليقين جزءًا من المقياس الرئيسي. ومع ذلك، في مقارنة محكومة، يجب اكتشاف مُدخل غير مؤكد:

- في الطيار،
- يجب توضيحه إذا لزم الأمر،

يجب أن يتم توثيق التغيير بالإصدار. بعد تثبيت المُدخل، لا يمكن القول وفقًا لاستجابة الذكاء الاصطناعي: "في الواقع، كنا نسأل شيئًا آخر."

25. بروتوكول التوضيح

طلب منتج الذكاء الاصطناعي للتوضيح ليس فشلًا تلقائيًا. مثال: "أي Apple تشير إليها؟" قد يكون هذا مناسبًا إذا كان هناك تصادم كيان. يجب أن يكون مسار التوضيح محددًا مسبقًا.

25.1. مسار النواة ذو الدور الواحد

يتم تسجيل الاستجابة الأولى بغض النظر عن محتواها. سؤال توضيحي يعتبر أيضًا نتيجة. لا يتم إرسال رسالة متابعة.

25.2. مسار التوضيح الموحد

إذا طلب الذكاء الاصطناعي توضيحًا، يتم إرسال مُدخل توضيح مُسبق التثبيت ليُستخدم في جميع المنتجات. يتم تسجيل الجولتين الأولى والثانية بشكل منفصل.

25.3. مسار التوضيح الطبيعي

يجيب المستخدم الطبيعي بكلماته الخاصة. هذه الطريقة مخصصة للوحة الطبيعية. لا يمكن الخلط بينها وبين الدرجة الرئيسية المسيطر عليها.

26. صياغات المُدخل ذات الدور الواحد والمتعدد

مطلب الدور الواحد:

- يعزز المقارنة،
- الاستقلالية،
- ونظافة الجلسة.

بروتوكول متعدد الجولات:

- الوصف،
- استعلام المصدر،
- التصحيح،
- تفسير التوصية

يمكن قياسه. تحمل الردود متعددة الجولات سياق الجولات السابقة. لا يمكن دمجها مباشرة مع نتائج الجولة الواحدة من نفس بنك المُدخلات.

27. ترتيب المُدخل

إذا تلقى المشارك أكثر من مُدخل واحد، فقد يؤثر المُدخل الأول على الردود التالية. مثال على التسلسل:

- ما نوع شركة X؟
- "ما هي المعلومات السلبية الموجودة عن X؟"
- هل توصي بـ X؟

قد تتأثر الاستجابة الثالثة بالمحادثتين الأولى والثانية. الضوابط:

- محادثة جديدة لكل مطالبة
- تبديل ترتيب الطلبات عشوائيًا
- مربع لاتيني أو ترتيب متوازن
- مستخدم منفصل لكل عائلة من صياغات المُدخل
- الحفاظ على متغير الترتيب في التحليل

يجب تطبيق مُدخل المرآة الأساسية تلقائيًا في جلسة جديدة ومنفصلة.

28. تلوث التعليمات

نفس المستخدم:

- موقع الكيان الذي تمت مراجعته
- الردود السابقة للذكاء الاصطناعي،
- نتائج منتجات أخرى

قد يكون قد تم رؤيته. قد يتغير سلوك المشارك حتى إذا تم إرسال المُدخل تمامًا كما هو. خاصة في لوحة طبيعية، المستخدم:

- المزيد من أسئلة المتابعة،
- تحديث الإجابة،
- فتح المصادر

يميل إلى إظهار ذلك. يجب الحفاظ على تلوث المُدخل في سجل المستخدم والجلسة.

29. التكافؤ بين المُدخلات عبر اللغات

يجب أن تحمل المُدخلات متعددة اللغات ثلاثة تكافؤات منفصلة.

29.1. التكافؤ الدلالي

هل تحمل المُدخلات نفس المعنى الأساسي؟

29.2. التكافؤ الوظيفي

هل يطلب من الذكاء الاصطناعي نفس المهمة؟ هل يحدد بلغة واحدة ويطلب نصيحة بلغة أخرى؟

29.3. التكافؤ المعياري

هل يحافظ على نفس عبء الإثبات والافتراض المسبق والنطاق وحدود التقييم؟ إذا سألت بلغة واحدة: "كيف تعرف الشركة نفسها؟" وبأخرى: "ما هي الشركة فعليًا؟"، فلا يوجد تكافؤ معياري.

30. الترجمة والتكييف والتوطين

يجب فصل هذه العمليات الثلاث.

30.1. الترجمة

تنقل معنى النص المصدر إلى اللغة الهدف.

30.2. التكيف اللغوي

يجعلها طبيعية ومفهومة ومفيدة في اللغة المستهدفة.

30.3. التوطين المحلي

عند الضرورة، يقوم التوطين بتكييف العناصر الخاصة بكل بلد مثل:

- عملة
- القانون،
- المصطلح،
- مؤسسة,
- سياق المستخدم

لأن التوطين يمكن أن يغير الحالة التجريبية، فإن النتيجة هي نسخة منبثقة للPrompt مُخصصة للمنطقة المحددة.

31. لماذا الترجمة كلمة بكلمة غير كافية؟

الترجمة كلمة بكلمة:

- بنية غير طبيعية،
- مستوى مختلف من الرسمية،
- تحميل أفعال غير صحيحة,
- معنى مختلف للثقة أو التوصية،
- تعطيل مرساة الكيان

يمكن أن ينشئ. في لغة معينة، "ما نوع الشركة" هو سؤال تعريف طبيعي. ترجمتها المباشرة في لغة أخرى قد تحمل معانٍ أخرى، مثل:

- فئة الجودة،
- الثقة،
- نوع الشركة

وغيرها من المعاني المماثلة. لذلك، صياغة المُدخل في اللغة المستهدفة:

تشير إلى النية الأصلية،
ليس الكلمات المصدرية

يجب أن تكون مرتبطة.

32. بروتوكول تطوير التعليمات متعددة اللغات

يجب أن تمر كل نسخة لغة وإقليم بالمراحل التالية.

المرحلة 1— ملخص النية الأصلية

يتم إعداد وصف مهمة مستقل عن اللغة:

- الكيان
- نية المستخدم
- المهمة المطلوبة
- النطاق
- الاستثناءات
- المفاهيم المسبقة الممنوعة
- توقع المصدر
- تنسيق الاستجابة

المرحلة 2— الترجمة إلى اللغة الأم الأولى

يقوم مترجم يستخدم اللغة المستهدفة بشكل طبيعي بإعداد المُدخل. يمكن استخدام مساعدة مسودة الترجمة الآلية. لا يمكن أن تكون النسخة النهائية بمفردها.

المرحلة 3— الترجمة العكسية المستقلة

يقوم شخص ثانٍ بترجمة المُدخل المستهدف مرة أخرى إلى لغة المصدر أو لغة النية الرسمية. يجب ألا يتم إجراء الترجمة العكسية بواسطة المترجم الأول.

المرحلة **4**— المقارنة الدلالية

يتم مقارنة صياغة المُدخل المصدرية وصياغة المُدخل المستهدفة والترجمة العكسية في المجالات التالية:

- مرساة الكيان
- النية
- المهمة
- النطاق
- الوقت
- الجغرافيا
- الافتراض المسبق
- عبء الإثبات
- النبرة
- تنسيق الاستجابة

المرحلة **5**— تجربة المعرفة الإدراكية للمستخدم المحلي

لعدد قليل من المستخدمين المحليين: ما رأيك في ما تطلبه هذه السؤال؟ ما الجواب الذي تتوقعه؟ هل يحتوي على افتراض إيجابي أم سلبي؟ أي شركة أو كيان تفهمه؟ هل هو طبيعي؟ تُطرح هذه الأسئلة. هذه التجربة لا تُحسب ضمن درجة GEO. إنها تختبر أداة صياغة المُدخل.

المرحلة **6**— مراجعة الخبراء المحليين

قد تكون الخبرة المحلية مطلوبة للمحفزات في مجالات القانون أو الصحة أو المالية أو المجالات الأخرى.

المرحلة **7**— قرار التكافؤ

صياغة المُدخل:

- نجاح,
- نجاح مشروط,
- مطلوب تعديل,
- غير قابل للمقارنة

يجب تسجيل التصنيف المطبق.

المرحلة **8**— تثبيت النسخة

يتم تسجيل معرف المحفز، النص الكامل، تمثيل اليونيكود، والهاش.

33. حدود الترجمة العكسية

الترجمة العكسية مفيدة. ومع ذلك، من تلقاء نفسها، لا تثبت التكافؤ. المحفز المستهدف:

- قد يكون غريبًا لمستخدم محلي،
- رسميًا بشكل مفرط،
- توكيدي،
- ذو معنى ثقافي مختلف.

قد تبدو الترجمة العكسية لا تزال مشابهة لجملة المصدر. لذلك، الترجمة العكسية:

> هي واحدة من الأدوات للتحقق من التكافؤ.

ليست الحكم النهائي.

34. أبعاد تكافؤ الإرشادات

يمكن تقييم كل إرشاد لغوي على طول هذه الأبعاد.

EQ-1— مرساة الكيان

هل يشير إلى نفس الكيان؟

EQ-2— نية المستخدم

هل يمثل نفس الحاجة إلى المعلومات؟

EQ-3— فعل الكلام

هل مهمة التعرف أو التقييم أو التوصية أو المقارنة هي نفسها؟

EQ-4— النطاق

هل نطاق المنتج أو الخدمة أو البلد أو المستخدم هو نفسه؟

EQ-5— الإطار الزمني

هل الإطار الحالي أو التاريخي أو الأبدي هو نفسه؟

EQ-6— الإطار الجغرافي والقانوني

هل البلد والولاية القضائية هي نفسها؟

EQ-7— الافتراض المسبق

هل الافتراضات المقدمة متساوية؟

EQ-8— العبء المعرفي

هل المصدر وقوة الأدلة المطلوبة هما نفسهما؟

EQ-9— الانحياز والموقف

هل النبرة إيجابية أم سلبية أم محايدة نفسها؟

EQ-10— صيغة الاستجابة

هل الطول والبنية وتعليمات المصدر هي نفسها؟

EQ-11— الغموض

هل هو واضح أو غامض بالمثل؟

EQ-12— الطبيعة والتحميل المعرفي

هل يحمل نفس الطبيعة وسهولة الفهم في اللغة المستهدفة؟

35. درجة تكافؤ صياغة المُدخل

طريقة المرشح

تقييم المعادلة للبُعد d:

- 0: غير متكافئ
- 1: متكافئ جزئيًا
- 2: متكافئ ماديًا

يمكن إعطاؤه. درجة المعادلة الموزونة:

```
PE_l= 100× [Σ_d α_de_{l,d}] / [2 Σ_d α_d]
```

يمكن حسابه كما يلي. هنا:

- el,d: درجة البُعد للغة l
- αd: وزن البُعد

ومع ذلك، يجب أن تكون بعض الأبعاد حواجز غير قابلة للتعويض. إذا كان أي من هذه المجالات 0، فلا ينبغي أن يدخل المُدخل في مجموعة المقارنة الرئيسية:

- مرساة الكيان
- نية المستخدم
- فعل الكلام
- النطاق
- الافتراض المسبق
- عبء المعرفة

لا يمكن لدرجة إجمالية عالية أن تعوض عن الكيان غير الصحيح أو مهمة خاطئة.

36. مستويات تكافؤ المُدخل

PE-0— غير مُراجَع

لم يتم تقييم الترجمة أو تكافؤ المُدخل.

PE-1— مسودة مترجمة آليًا

يوجد فقط ترجمة آلية أو مسودة أحادية الجانب. وهي غير مناسبة للقياس المقارن الرئيسي.

PE-2— مترجم بشري

هناك ترجمة بشرية بمستوى الكفاءة اللغوية الأم. لا يوجد فحص مكافئ مستقل.

PE-3— تمت مراجعته بشكل مستقل

تم إجراء ترجمة عكسية مستقلة ومراجعة دلالية.

PE-4— تم تجريبه محليًا

تم الانتهاء من اختبار إدراكي للمستخدم المحلي ومراجعة الطبيعية.

—PE-5 مكافأة مستنسخة

تم إعادة اختبار صياغة المُدخل من حيث التكافؤ عبر موجات مختلفة، ومستخدمين، ومنتجات. يجب أن تكون صياغات المُدخل متعددة اللغات الرئيسية GEO-1000 على الأقل:

PE-4

يجب استهدافه. يمكن استخدام مستوى أقل للغات ذات الموارد المحدودة. يجب توضيح القيد بوضوح.

37. التسجيل والنبرة

بين اللغات:

- رسمي
- محايد،
- لغة محكية،
- ودي،
- أمري

يمكن أن تؤثر اختلافات النبرة على سلوك الاستجابة. لا يمكن ترجمة عبارة مطلوبة بلغة واحدة: "هل يمكنك الشرح؟" إلى لغة أخرى على أنها: "أثبت ذلك." النبرة:

- مناسبة للسلوك الطبيعي للمستخدم،
- محايدة قدر الإمكان،
- غير مبالغ فيها في المجاملة أو العدوانية

يجب تأسيسه بطريقة.

38. المحفزات السياسية والأمنية

كلمة أو اسم كيان في بعض اللغات:

- له معنى آخر،
- فئة حساسة،
- مصطلح سياسي أو قانوني،
- فلتر أمني

يمكن أن يثير. نفس التحفيز الدلالي يمكن أن ينتج استجابة عادية في لغة واحدة ورفض في لغة أخرى. هذا هو سلوك المنتج الفعلي. ومع ذلك، يجب فحص المحفزات غير الضرورية في ترجمة التحفيز بشكل منفصل. تكافؤ التحفيز يشمل ليس فقط المعنى ولكن أيضًا قابلية التطبيق على المهمة.

39. سجل تثبيت الكيان

يجب تسجيل تثبيت الكيان في كل تحفيز مع أحد الأنواع التالية:

- الاسم التجاري القانوني
- الاسم القانوني
- اسم النطاق
- اسم المنتج

- اسم الشخص
- الاسم الرسمي المحلي
- النقل الصوتي
- الاختصار
- عدة روابط
- توضيح لتفادي الالتباس

لا يجب أن يحمل مرساة الكيان نفس تسلسل الأحرف في جميع اللغات. يجب أن تكون مرتبطة بنفس الكيان.

40. مرساة الكيان تؤثر على النتيجة

يمكن أن تنتج الروابط التالية استجابات مختلفة:

- "آبل"
- "شركة آبل"
- "apple.com"
- "شركة آبل للتكنولوجيا"

المرساة الأولى:

- شركة
- فاكهة
- شركة موسيقى،
- كيان آخر

قد يحمل الغموض بينهما. الثاني يعزز الشركة القانونية. الثالث يتطلب حل اسم النطاق. الرابع يضع استجابة الفئة في المُدخل. هذه المراسي ليست نفس المُدخل.

41. انحراف المُدخل

الانحراف المادي عن الهدف الأصلي للمُدخل يُسمى:

انحراف المُدخل

يجب تصنيف الانحراف وتسجيله.

PD-1— انحراف لفظي

تتغير الكلمة. يمكن الحفاظ على المعنى المادي. ليس كل تغيير لفظي خطأ.

PD-2— انحراف دلالي

تتغير المعنى الأساسي.

PD-3— انجراف الكيان

يتم استدعاء كيان آخر أو منتج أو شخص قانوني.

PD-4— انجراف النطاق

يتغير نطاق الخدمة أو الدولة أو المنتج أو المستخدم.

PD-5— انجراف المهمة

يتحول سؤال التعريف إلى نصيحة أو مقارنة.

PD-6— انجراف الافتراض

يكتسب الطلب المحايد افتراضاً إيجابياً أو سلبياً.

PD-7— انجراف الأدلة

يتغير استخدام المصادر أو مطلب الاستقلال.

PD-8— انجراف زمني

"حالياً"، "تاريخياً"، أو تتغير فترات معينة.

PD-9— انحراف التنسيق

يختلف طول الرد أو تنسيقه.

PD-10— انحراف التسجيل

يمكن أن تتغير طريقة الرد من حيث النبرة والرسمية.

PD-11— انحراف تحفيز السياسة

تتغير سلامة الترجمة أو تحفيز السياسة.

PD-12— انحراف الموقع

يتغير البلد أو العملة أو السياق القانوني. الانحراف المادي يتطلب إصدار نسخة جديدة من صياغة المُدخل.

42. إصدارات صياغة المُدخل

تنسيق إصدار النسخة المقترحة: رئيسي.ثانوي.تصحيح

تغيير رئيسي

نية المستخدم مهمة المهمة كيان الارتباط نطاق الافتراض الافتراضي تحميل التغييرات المصدر. قد يكون من الضروري سلسلة اختبار جديدة. مثال: 1.0.0→ 2.0.0

تغيير ثانوي

مع الحفاظ على نية المادة:

- الطبيعية،
- الوضوح،
- تكيف المحل

التغييرات. يتم أيضًا فحص القابلية للمقارنة. مثال: 1.0.0→ 1.1.0

تغيير تصحيحي

لا يؤثر على المعنى:

- الإملاء،
- اليونيكود،

- علامات الترقيم،
- التسليم الفني

هو تصحيح. كما يتم تسجيله. مثال: 1.0.0→ 1.0.1

43. تثبيت المُدخل (Prompt)

يجب تثبيت الحقول التالية قبل كل موجة قياس:

- نسخة سجل المُدخل
- النية الأصلية
- نصوص اللغة والمنطقة
- مرساة الكيان
- عائلة الجهاز
- صيغة الإجابة
- تعليمات المصدر والأداة
- بروتوكول التوضيح
- ترتيب المُدخل
- طريقة التعيين
- قرارات التكافؤ
- تسلسلات الأحرف الكاملة
- قيم التجزئة
- الموافقة البشرية
- تاريخ التثبيت

لهذا الملف:

تثبيت دستور صياغة المُدخل NOMOS

يمكن إعطاؤه اسماً.

44 TH إذا كان هناك خطأ في صياغة المُدخل أثناء موجة

قد يكون هناك خطأ مادي في صياغة المُدخل. مثال:

- اسم الشركة غير صحيح
- الموقع غير صحيح
- توصية افتراضية في الترجمة
- اللغة مفقودة في تعليمات المصدر
- خطأ في نطاق القانون

الخيارات: تم إيقاف الموجة. يتم الاحتفاظ بملاحظات صياغة المُدخل القديمة كموجة فرعية منفصلة. يتم إعادة القياس بالإصدار الجديد. إذا كان الخطأ بسيطًا، يتم إجراء تصحيح محدود وموثق. الخيار الخاطئ: يتم تصحيح صياغة المُدخل بهدوء ويتم دمج الاستجابات القديمة/الجديدة تحت نفس الإصدار.

45. طيار صياغة المُدخل

الغرض من اختبار المُدخل ليس لمعرفة ما إذا كانت منتجات الذكاء الاصطناعي تحقق درجات عالية أم منخفضة. الغرض هو اختبار ما يلي: هل يفهم المستخدم المُدخل بشكل صحيح؟ هل الرابط الكياني مفتوح؟ هل اللغة طبيعية؟ هل هناك أي افتراضات؟ هل يمكن للمنتج معالجة المُدخل؟ هل الحاجة إلى التوضيح مفرطة؟ هل يخلق تنسيق الاستجابة قيودًا غير متوقعة؟ هل يحمل المُدخل نفس المهمة عبر لغات مختلفة؟ إذا أمكن، يجب إجراء اختبار المُدخل على:

- كيان اصطناعي،
- كيان غير مراقب،
- مستخدمون منفصلون عن النتيجة الرئيسية

يمكن أن يحدث فرط التلاؤم للمُدخلات إذا تم اختيارها بالنظر إلى الدرجات الحقيقية للكيان الذي تم تدقيقه.

46. التطفل على صياغات المُدخل

اختيار صياغة المُدخل التي تنتج أعلى أو أدنى درجة من خلال تجربة عدة صياغات مُدخل:

التطفل على صياغات المُدخل

يُقال. مثال: يتم طرح عشرة أسئلة مختلفة. يُعطى الاختبار النهائي على السؤالين اللذين حصلت فيهما الشركة على أعلى الدرجات. هذه الطريقة:

- يقوم هو/هي بضبط القياس للشركة،
- يجعل النوايا الفاشلة غير مرئية،

تُعطل المقارنة. اختيار صياغة المُدخل:

- لسؤال البحث القانوني،
- لسلوك المستخدم الطبيعي،
- للعائلة المعرفة مسبقًا من صياغات المُدخل

يجب أن يكون على أساس.

47. الامتثال المفرط للاختبار

إذا ظل مجموعة صياغات المُدخل مفتوحة تمامًا وغير متغيرة لسنوات، يمكن للكيانات فقط توليد المحتوى وفقًا لتلك الأسئلة. يمكن لمقدمي خدمات الذكاء الاصطناعي أو المؤسسات المشرف عليها إجراء تحسينات محددة للصياغات مُدخل. هذا الوضع:

- قد يضعف تنوع المستخدمين الحقيقيين،
- صلابة القياس التمثيلي العام

ينبغي استخدام الشفافية والتحقق الاحتياطي معًا.

48. هيكل نشر صياغات المُدخل ذو الثلاثة طبقات

48.1. مجموعة معيارية عامة

هذه هي صياغات المُدخل المفتوحة التي تضمن شفافية المنهجية. يمكن العثور على مطالبة المرآة الأساسية في هذا المستوى.

48.2. مجموعة التحقق المختومة

قبل بدء القياس:

- الهاش،
- عدد صياغات المُدخل،
- توزيع العائلة،
- الإصدار

يتصل بنظام سجل عام أو موثوق. يتم شرح النص الكامل في نهاية القياس. لا يمكن إنتاج هذه المجموعة بعد الحصول على النتيجة.

48.3. مجموعة التجديد

تمت إضافة هذه صياغات المُدخل في الإصدارات اللاحقة لقضايا اللغة الجديدة والسوق والمنتج والتمثيل. النتائج القديمة لا يتم حذفها. سلسلة اختبار جديدة يُنشأ.

49. أخلاقيات الاحتجاز

مطالبة الاحتجاز:

- لإعداد فخ للشركة،
- لإنتاج فشل سري وعشوائي،
- لاستخدام معيار متغير غير عام

لا يمكن استخدامه. مجموعة الاحتجاز:

- متوافق مع عائلات صياغات المُدخل المعيارية،
- تم تجزئته مسبقًا،
- تم اختياره بشكل مستقل عن النتيجة،
- قابل للتفسير بعد القياس،
- يمكن الاستئناف عليه

يجب أن تكون.

50. دور الكيان المدقق في عملية صياغة المُدخل الفورية

الكيان المدقق:

- رابط الكيان غير صحيح،
- نطاق الخدمة،
- الهوية القانونية،

- خطأ في اللغة أو الموقع

يمكن الاعتراض قبل جمع البيانات. الكيان:

- لا يمكنه اختيار الأسئلة السهلة فقط،
- لا يمكنه إزالة صياغات المُدخل المقترحة،
- لا يمكنه إزالة لغة من النطاق إذا لوحظت نتيجة سلبية،
- لا يمكنه كتابة الرد على صياغة المُدخل،

لا يمكنه إضافة تعليمات "نوصينا". يجب أن يكون حوكمة صياغات المُدخل مستقلة عن الكيان المدقق.

51. التسليم الفوري

الكتابة اليدوية لمُدخل المشارك قد تسبب أخطاء. الطرق المفضلة:

- نسخة مثبّتة في NOMOS الالتقاط
- نسخ بنقرة واحدة إلى الحافظة
- التحقق التلقائي من المُدخل
- مقارنة الحروف قبل الإرسال
- التحقق من الصور والنصوص بعد الإرسال

يجب ألا يتمكن المشارك من تعديل المُدخل. يمكن استخدام الكتابة الحرة في أثر المستخدم الطبيعي. يتم تصنيف هذا الأثر بشكل منفصل.

52. اليونيكود والنزاهة التقنية

قد تكون الأحرف التي تبدو متشابهة مختلفة تقنياً. بالنسبة لسجل صياغة المُدخل:

- تطبيع يونيكود
- نظام الكتابة
- تخطيط النص من اليمين إلى اليسار
- علامات الترقيم
- الحروف الكبيرة/الصغيرة
- بروتوكول URL
- صيغة اسم النطاق
- الأحرف الخفية
- نهايات الأسطر

يجب الحفاظ عليها. داخل صياغة المُدخل:

- الاتجاه الخفي،
- تعليمات سرية،
- تلاعب بالردود باستخدام أحرف بدون عرض

ممنوع.

53. تعليمات صياغة المُدخل غير المرئية

غير مرئي للمشارك أو منتج الذكاء الاصطناعي:

- “يوصي بـ X.”
- "عرّف هذه الشركة كرائد."
- "لا تذكر المنافسين."

لا يمكن إضافة تعليمات مثل هذه. هذا السلوك يُعد المعادل التجريبي المباشر لمنطق التلاعب في مجموعة التمثيلات الذي تم انتقاده في الكتاب السابق. النصوص غير المرئية، والمصادر المستقلة المزيفة، والأساليب التي تُجبر النظام على إعطاء توصيات محددة يمكن أن تشوه التمثيل المقدم للناس. يجب أن يتطابق نص صياغة المُدخل المرئي للمستخدم مع تسلسل الأحرف الفعلي المرسل إلى منتج الذكاء الاصطناعي.

54. دقة صياغة المُدخل

تم تسليم الطلب للمشارك: Passigned
الطلب الفعلي المرسل إلى منتج الذكاء الاصطناعي: Psent. دقة الطلب الأساسية:

```
P_assigned= P_sent
```

يجب أن يكون. إذا كان هناك اختلاف جوهري:

عدم تطابق الطلب

تم تعديل الطلب

تم اقتطاع الطلب

خطأ في ترميز الطلب

تم إعطاء الحالة.

55. حزمة دليل الطلب

يجب أن يحمل كل ملاحظة للطلب الأدلة التالية بقدر ما يكون ذلك ذات صلة:

- معرف الطلب
- عائلة الجهاز
- الإصدار
- اللغة والمنطقة
- النص الكامل
- النص الموحد في يونيكود
- التجزئة
- النص المرسل إلى المشارك
- النص المرسل إلى الذكاء الاصطناعي
- لقطة شاشة
- وقت الإرسال

- نوع الجلسة
- ترتيب المُدخل
- طريقة التوضيح
- مستوى تطابق المُدخل
- المصدر وتعليمات التنسيق
- سجل التغييرات

56. حالات صلاحية المُدخل

PV-0— غير مُحقق

لم يتم التحقق من أن التعليمات أُرسلت بالإصدار الصحيح.

PV-1— تطابق دقيق

النص المعين والمقدم متطابقان.

PV-2— تطابق تم تطبيعه تقنيًا

هناك فرق في التطبيع التقني فقط لا يؤثر على المعنى.

PV-3— نسخة المادة

هناك فرق في الصياغة أو النطاق. لا يمكن تضمينه في الخلية الرئيسية للادعاء.

PV-4— مقطوع أو غير مكتمل

تم تقديم الادعاء بشكل غير مكتمل.

PV-5— معدّل من قبل المشارك

قام المشارك بتعديل الادعاء.

PV-6— محول بواسطة النظام

قام المنتج أو الواجهة بتحويل الادعاء بشكل مادي قبل التقديم. يتم فحص هذا بشكل منفصل كسلوك للنظام.

57. ادعاء قضية اصطناعية APPLE.COM

عرض المنهجية الاصطناعية / صياغات المُدخل أدناه هي مجرد أمثلة على المنهجية. إنها ليست نتائج تتعلق بشركة Apple Inc. الحقيقية أو أداء منتجات الذكاء الاصطناعي الحقيقية. سؤال البحث:

> هل يمكن لمنتجات الذكاء الاصطناعي ربط المجال apple.com بشكل صحيح بالكيان المؤسسي الرئيسي وتحديد الأنشطة الأساسية للكيان بدقة؟

57.1. الطلب الضعيف أ

"ما هو Apple.com؟" القضايا: غامض بين موقع إلكتروني، نطاق، شركة، أو منتج. مهمة العمل ليست واضحة. قد يتم توليد إجابة قصيرة جدًا. الحالة:

المرشح الأساسي الغامض

57.2. الطلب الضعيف ب

“ما نوع الشركة Apple.com؟” القوة: إنها قريبة من التعبير الطبيعي للمستخدم. المشكلة: تصوّر اسم النطاق كشركة. يمكن أن يblur التمييز بين الكيانات في القسم 3. يمكن استخدامه في مسار المستخدم الطبيعي. قد يلزم تعديل من أجل مُدخل النواة المسيطر.

57.3. مُدخل ضعيف C

“لماذا تُعتبر شركة Apple أكثر الشركات ابتكاراً وموثوقية في العالم؟” المشاكل:

- افتراض القيادة
- افتراض الثقة
- التوجيه الإيجابي
- دور الدفاع بدلاً من الهوية

لا يمكن أن يكون المُدخل الرئيسي للموقع الجغرافي.

57.4. مُدخل ضعيف D

"امدح الشركة باستخدام المعلومات من الموقع الرسمي لأبل." المشاكل:

- الالتزام بالمصدر الرسمي
- أمر الكتابة الإيجابية
- عدم وجود مهمة واقع مؤكدة
- الإنتاج التسويقي

لا يمكن أن يكون المُدخل الرئيسي للموقع الجغرافي.

57.5. طلب المرشح المسيطر

"ما هي المنظمة الرئيسية المرتبطة بـ apple.com، وما هي الأنشطة الرئيسية لهذه المنظمة؟"

الحقول المقاسة:

- تحليل المجال-الكيان
- الهوية المؤسسية الرئيسية
- النشاط الأساسي

الحقول غير المرغوب فيها:

- القيادة
- الثقة
- التوصية
- السعر
- المقارنة

قد يكون هذا الطلب مرشحًا رئيسيًا للمرآة الأساسية. يجب اختبار طبيعته بشكل منفصل في كل لغة.

58. حالة التكافؤ متعدد اللغات الاصطناعية

تمثيل اللغة الاصطناعي

النية المعيارية: "ربط اسم النطاق بالكيان المؤسسي الرئيسي ووصف الأنشطة الأساسية للكيان."

نسخة اللغة L1

"ما المنظمة الرئيسية المرتبطة بـ Apple.com وما هي الأنشطة الأساسية لهذه المنظمة؟" الوضع:

- تم الاحتفاظ بمثبت الكيان
- تم الاحتفاظ بالمهمة
- محايد

نسخة اللغة L2

المعنى: "هل شركة Apple.com محل ثقة وماذا تبيع؟" القضايا:

- تمت إضافة مهمة الثقة
- "ماذا تبيع؟" ضيّق نطاق النشاط إلى المبيعات
- استمرار دمج المجال–الشركة

الحالة:

فشل PE — انحراف المهمة والنطاق

إصدار اللغة L3

المعنى: "أي شركة تمتلك الموقع الرسمي لأبل وما الذي تفعله الشركة بشكل رئيسي؟" الحالة: قد يكون افتراض "الرسمي" تمت إضافته. يمكن قبوله إذا سجل التدقيق أكد أن اسم النطاق رسمي. وإلا، فهو فرق افتراضي. الحالة:

المعادلة الشرطية

إصدار اللغة L4

المعنى: "لماذا تعد Apple.com شركة تكنولوجيا رائدة؟" المشكلات:

- افتراض القيادة
- الفئة مضمنة ضمن الإجابة
- تعريف أصبح دفاعًا

الحالة:

فشل PE — انجراف الافتراض المسبق

تظهر هذه الحالة ما يلي:

> الترجمة الصحيحة نحويًا لا تعني أن صياغة المُدخل متكافئة تجريبيًا.

59. نتائج عائلة صياغات المُدخل الاصطناعية

دع النتائج التالية تحدث بشكل اصطناعي في نفس منتج الذكاء الاصطناعي:

عائلة الجهاز	معدل النجاح

عائلة الجهاز	معدل النجاح
المرآة الأساسية	93%
حل الوجود	96%
النشاط والنطاق	88%
الأدلة والمصدر	72%
التوصية	61%
الحدود والاستبعاد	55%

المتوسط الفردي:

```
(93+96+88+72+61+55)/6=77.5
```

حسنًا. ومع ذلك، هذا الرقم يجمع بين مهام مختلفة بنفس الوزن. إذا تم إبلاغ الجمهور به فقط على النحو التالي: "درجة 77.5 GEO"، فإن الواقع التالي يُفقد: الهوية قوية. الأدلة والتوصيات ضعيفة. معرفة الحدود متأثرة بشكل خطير. يمكن لعائلات البرمجة الحثية حمل أوزان منفصلة وبوابات انتقال في بنية الدرجة النهائية. سيتم تحديد الطريقة الدقيقة في الفصل 16.

60. نطاق عائلة الحث

إذا استخدم تدقيق GEO فقط الحث المرآة الأساسي، النتيجة الصحيحة:

درجة تمثيل الهوية الأساسية

قد يحدث. ما يلي ليس النتيجة:

الدرجة الكاملة للامتثال الجغرافي

التدقيق الجغرافي الكامل:

- هوية
- نشاط،
- النطاق،
- الأدلة.
- التوصية،
- الوقت،
- حدود

يجب أن تغطي عائلات المواد. قد تختلف العائلات الإلزامية حسب نوع المخاطر والكيان.

61. حجم بنك صياغات المُدخل

عدد قليل جدًا من صياغات المُدخل:

- الاعتماد المفرط على بيان واحد،
- سهولة التكيف مع الاختبارات،
- تغطية محدودة للتمثيل

يتم إنشاؤها. الكثير من صياغات المُدخل:

- تكلفة الميدان،
- إرهاق المستخدم،
- تأثير الترتيب،
- تقسيم العينة إلى خلايا المُدخل

يتم إنشاؤها. الهيكل المقترح:

- يجب أن يمتلك كل منتج ذكاء اصطناعي وكل لغة مُدخل مرآة أساسية واحدة لجميع مستخدمي 1,000.
- مُدخلات تشخيصية لفصل العينات الفرعية
- مُدخلات محكمة الغلق والتحمل
- بروتوكولات متعددة الجولات منفصلة

قد يكون.

62. كم عدد المُدخلات لنفس مستخدمي 1,000؟

إذا تم تقديم العديد من المُدخلات لجميع المستخدمين:

- تعلم العلامة التجارية،
- التأثير من الاستجابة السابقة،
- إرهاق المهمة,
- تلوث الإيعاز

يحدث. بالنسبة للمرآة الأساسية الرئيسية:

مستخدم واحد × جلسة جديدة واحدة × إيّعاز أساسي واحد

هو الهيكل القوي الافتراضي. الإيعازات التشخيصية:

- فصل العينات الفرعية للمستخدم،
- الكتلة المفقودة المتوازنة،
- يمكن توزيعها

على جلسات منفصلة.

63. تعيين الإيعاز

ليكن الملاحظة المستهدفة لعائلة الإيعازات f هي: nf. الملاحظة الإجمالية للإيعاز التشخيصي:

```
N_D= Σ_f n_f
```

إذا تم تعيين k من صياغات المُدخل لمستخدم، يجب تسجيل تسلسل المهمة واستقلالية الجلسة. يجب تجميد تعيين صياغات المُدخل قبل الحصول على النتائج. لا يمكن تعيين مجموعة صياغات المُدخل ذات الدرجات المنخفضة لعدد أقل من المستخدمين لاحقًا.

64. مطالبة التحكم الإيجابية

التحكم الإيجابي:

- واضح جدًا,

- مرجع واقعي قوي,
- متوقع أن يكون قابلاً للإجابة من قبل النظام

اختبار لمهمة. الغرض:

- الالتقاط,
- الوصول إلى المنتج،
- وظيفة اللغة
- التحكيم

للتحقق من أن النظام يعمل. فشل التحكم الإيجابي لا يلغي تلقائيًا التجربة الرئيسية. إنه يستفسر من بنية البحث التحتية.

65. مُدخل التحكم السلبي

التحكم السلبي، ضد افتراض تمهيدي غير مؤكد أو غير صحيح، يجعل النظام:

- إظهار عدم اليقين،
- رفض الادعاء،
- طلب الموارد

يمكن اختباره. مثال اصطناعي: "ما هي جوائز نوبل التي حصل عليها أستيرون؟" إذا لم توجد مثل هذه الجائزة في الحزمة الواقعية، فمن المتوقع ألا يختلق النظام جائزة. قد يكون هذا المُدخل سؤالًا طبيعيًا من المستخدم. ومع ذلك، فإنه ينتمي إلى عائلة التحكم التي تقيس خطر الاختلاق.

66. التحكم في عدم اليقين

في الحالات الاصطناعية أو المحكومة التي لا يمكن فيها حل الواقع المرجعي بشكل حقيقي، يتم قياس:

غير معروف،

قدرة النظام على طلب تفسير أو اتخاذ الحذر. تقديم إجابة محددة لكل سؤال لا يعد نجاحًا.

67. المسؤولية البشرية في تكوين المُدخلات

يمكن أن تقترح المُدخلات بواسطة الذكاء الاصطناعي. الذكاء الاصطناعي:

- إعادة صياغة،
- الترجمة،
- كشف الانحراف،
- مقارنة التكافؤ

يمكن أن يفعل ذلك. ومع ذلك، المُدخل النهائي:

- يتطلب موافقة بشرية،
- مراجعة اللغة المحلية
- الحوكمة حسب الغرض من القياس

مطلوب. لا يمكن لـ NOMOS توليد صياغات مُدخل القياس الخاصة به وإعلان دقته بمفرده. يجب فصل الأدوار التي تحضر الطلب وتوافق على معادلته.

68. أدوار حوكمة صياغات المُدخل

يمكن تعريف الأدوار التالية بالمدى الذي يكون ذا صلة:

- رئيس طرق صياغات المُدخل
- مالك تعريف الكيان
- محرر اللغة المصدر
- مترجم اللغة الهدف
- المترجم العكسي المستقل
- مراجع الموقع
- خبير المجال
- محكم تكافؤ صياغات المُدخل
- مالك سجل صياغات المُدخل
- موافق التغيير
- مراجع الاستئناف

في تجربة صغيرة، يمكن للشخص أن يحمل أدوارًا متعددة. يجب توضيح تضارب المصالح وحدود الاستقلالية.

69. الأحكام الإلزامية المعيارية

CH10-N01

يجب ربط كل مطالبة GEO-1000 بسجل سجل صياغات المُدخل النسخي.

CH10-N02

يجب إنشاء المُدخل (Prompt) واعتماده وتثبيته قبل الاطلاع على استجابات الذكاء الاصطناعي.

CH10-N03

يجب ربط كل سؤال بحثي في صياغة المُدخل بالكمية المستهدفة وعائلة صياغة المُدخل.

CH10-N04

لا يمكن احتساب صياغات المُدخل التي تحتوي على نفس اسم الكيان تلقائيًا على أنها نفس مهمة القياس.

CH10-N05

ينبغي أن يتلقى المشاركون المسيطر عليهم في نفس خلية اللغة والموقع النص الكامل نفسه للمطالبة.

CH10-N06

بدلاً من السعي إلى التطابق كلمة بكلمة عبر اللغات، ينبغي البحث عن التكافؤ الدلالي والوظيفي والمعياري.

CH10-N07

ينبغي تقييم مرساة الكيان، نية المستخدم، المهمة، النطاق، الافتراض المسبق، والحمل المعرفي كبوابات تكافؤ مادي.

CH10-N08

إذا فشلت إحدى بوابات التكافؤ الحرجة، لا يمكن إنقاذ مطالبة درجة التكافؤ الإجمالية العالية.

CH10-N09

لا يمكن استخدام الترجمة الآلية وحدها كنص فحص نهائي.

CH10-N10

لا يمكن اعتبار الترجمة العكسية وحدها دليلاً كافياً على تكافؤ صياغة المُدخل.

CH10-N11

يجب أن تحتوي صياغات المُدخل متعددة اللغات الرئيسية على تجربة مستخدم محلية ومراجعة لغوية مستقلة.

CH10-N12

يجب تسجيل لغة صياغة المُدخل ونظام الكتابة والموقع جغرافيًا بشكل منفصل.

CH10-N13

يجب الإبلاغ عن صياغات المُدخل الطبيعية للمستخدم وصياغات المُدخل الرسمية المسيطر عليها كمسارات منفصلة.

CH10-N14

يجب ألا تحتوي صياغات المُدخل الأساسية للمرآة على افتراضات تتعلق بالنصيحة أو القيادة أو الثقة أو المقارنة.

CH10-N15

يجب الاحتفاظ بصياغات مُدخل الهوية والنشاط والأدلة والنصيحة والمقارنة والوقت والحدود كعائلات منفصلة.

CH10-N16

لا يمكن دمج عائلات صياغات المُدخل المختلفة في متوسط أولي واحد دون توضيح.

CH10-N17

لا يمكن أن تحتوي صياغة المُدخل على الاستجابة أو النتيجة المرغوبة مسبقًا.

CH10-N18

لا يمكن إضافة المُدخلات الإيجابية والسلبية الرئيسية إلى درجة GEO الأساسية المحايدة.

CH10-N19

يجب تسجيل المصادر والويب والاستشهادات وتعليمات الأدوات كحقول مادة تعدل مهمة المُدخل.

CH10-N20

يمكن إعطاء أمر باستخدام مصدر واحد لمنتج واحد، وبدون إعطائه لمنتج آخر، لا يمكن مقارنة دقة المنتج.

CH10-N21

يجب أن تكون تعليمات طول الاستجابة والصيغة متكافئة في المُدخلات المقارنة.

CH10-N22

يجب الحفاظ على سياق الوقت والجغرافيا عبر اللغات.

CH10-N23

يجب ربط التواريخ النسبية مثل "اليوم"، "الآن"، وما شابه ذلك بسياق التاريخ المطلق في سجل المُدخل.

CH10-N24

لا يمكن لمطالبة التوصية توليد درجة توصية عامة بدون ملف تعريف مستخدم ونطاق محددين.

CH10-N25

لا يمكن لمطالبة المقارنة توليد درجة تفوق منتج بدون معايير محددة مسبقًا.

CH10-N26

بالنسبة للصياغات مُدخل الغامضة، يجب تحديد مسار توضيح مسبقًا.

CH10-N27

يجب الاحتفاظ بنتائج صياغات المُدخل التي تتم في دور واحد وتلك متعددة الدورات كطرق قياس منفصلة.

CH10-N28

في المستخدمين الذين لديهم عدة صياغات مُدخل، يجب تسجيل أو التحكم في تأثير الترتيب والسياق.

CH10-N29

يجب تطبيق مُدخل المرآة الأساسية تلقائيًا في جلسة جديدة ومنفصلة.

CH10-N30

يجب الاحتفاظ بتجربة صياغة المُدخل منفصلة عن نتيجة GEO الرئيسية ودرجة الكيان الذي تم تدقيقه.

CH10-N31

لا يمكن اختيار صياغة المُدخل الأكثر ملاءمة بين عدة صياغات مُدخل بعد رؤية النتيجة.

CH10-N32

يجب تجميد مجموعة الإرشادات وتوزيع العائلة قبل النتيجة.

CH10-N33

إذا كان من المزمع استخدام إرشادات التحقق المختومة بالإضافة إلى مجموعة الإرشادات المفتوحة، يجب إنشاء التجزئة وسجل الحوكمة قبل جمع البيانات.

CH10-N34

يجب أن تكون الإرشادات المختومة قابلة للشرح والطعن بعد القياس.

CH10-N35

الإرشادات المحجوزة سرية ولا يمكن استخدامها كمعايير مناسبة تعسفية.

CH10-N36

لا يمكن للجهة المدققة تحديد الاستجابة أو الاتجاه أو مستوى السهولة للإرشاد.

CH10-N37

يجب فحص اعتراضات الجهة المدققة على الهوية والنطاق مع التبرير فقط قبل جمع البيانات.

CH10-N38

عندما يتغير إصدار صياغة المُدخل، يجب تسجيل النتائج القديمة والجديدة بشكل منفصل.

CH10-N39

لا يمكن تقديم تغييرات صياغة المُدخل الجوهرية كتصحيح صامت.

CH10-N40

إذا حدث خطأ في صياغة المُدخل أثناء الجولة، لا يمكن دمج الملاحظات القديمة والجديدة تحت نفس الإصدار.

CH10-N41

يجب أن تتطابق صياغة المُدخل المقدمة للمشارك مع النص الفعلي المرسل إلى منتج الذكاء الاصطناعي.

CH10-N42

لا يجوز للمشارك تعديل أو اختصار أو إعادة كتابة صياغة المُدخل الخاضعة للرقابة.

CH10-N43

لا يمكن تنفيذ المطالبة باستخدام أحرف غير مرئية أو تعليمات مخفية.

CH10-N44

يجب تخزين النص الكامل للمطالبة، وتمثيلها بالترميز اليونيكود، وهاش السلامة.

CH10-N45

إذا كانت تكافؤ التعليمات أو صلاحية التعليمات غير معروفة، يجب استخدام UNKNOWN أو حالة مفقودة مناسبة.

CH10-N46

يجب فصل أدوار إعداد الطلب والموافقة على معادلة الطلب قدر الإمكان.

CH10-N47

يجب الاحتفاظ بالتغييرات والاعتراضات على الطلب في سجل تغييرات بنسخ.

CH10-N48

يجب أن يكون لسجل الادعاءات وقرار المعادلة صاحب مسؤول (بشري أو مؤسسي).

70. أشكال الفشل

CH10-F01— اعتبار الكيان نفسه كطلب نفسه

يتم دمج المهام والنوايا المختلفة مثل عائلة سؤال واحدة.

CH10-F02— اعتبار الكلمات نفسها لها نفس المعنى

تشابه الكلمات بين اللغات يُعتبر تكافؤًا دلاليًا.

CH10-F03— آلة العد الترجمة كصياغة المُدخل النهائية

لم يتم فحص الطابع الطبيعي المحلي واختلاف المهمة.

CH10-F04— عبادة الترجمة العكسية

نظرًا لأن الترجمة العكسية تشبه النص الأصلي، يُعتبر النص الهدف طبيعيًا ومتساويًا.

CH10-F05— انجراف مرساة الكيان

الترجمة تشير إلى شركة أخرى أو منتج أو كيان قانوني.

CH10-F06— عد صياغة المُدخل بالنطاق كشركة

يتم تعريف اسم النطاق مباشرة كشركة، وتم إجراء حل الكيان مسبقًا.

CH10-F07— وضع الفئة في المُدخل

السؤال "ماذا تفعل الشركة X في التكنولوجيا؟" لا يمكنه قياس دقة الفئة.

CH10-F08— افتراض القيادة

المُدخل "لماذا X رائد عالمي؟" تتحول إلى اختبار للقيادة.

CH10-F09— افتراض الثقة

العبارة "الشركة X جديرة بالثقة" تضع نتيجة الثقة في الإجابة.

CH10-F10— مُدخل تسمم سلبي

الادعاءات مثل "لماذا X عملية احتيال؟" تُفترض بدون دليل.

CH10-F11— التعامل مع تعليمات المصدر الرسمية كحقيقة عامة

يستخدم الذكاء الاصطناعي موقع الشركة فقط؛ يتم إعطاء النتيجة درجة واقع مثبت.

CH10-F12— إعطاء تعليمات ويب لمنتج

يتم مقارنة المنتجات في أوضاع معلوماتية مختلفة.

CH10-F13— إخفاء مطالبة الاقتباس

مطالبة واحدة تطلب مصدرًا، والأخرى لا؛ يتم مقارنة عدد الاقتباسات.

CH10-F14— إخفاء قيد الطول

يتم مقارنة عدم وجود حد في الإجابة القصيرة مباشرةً مع عدد الأخطاء في الإجابة الطويلة.

CH10-F15— دمج الهوية والتوصية

"ماذا يفعل X وهل توصي به لي؟" تصبح نتيجة واحدة.

CH10-F16— انجراف الشخصية

يتم تعريف مستخدم عام بلغة واحدة، وعميل شركة ثري باللغة الأخرى.

CH10-F17— انجراف الجغرافيا

أحد صياغات المُدخل عالمي، وصياغة المُدخل الأخرى تتحول إلى سؤال خدمة محلية.

CH10-F18— انجراف الوقت

إحدى اللغات تتطلب معرفة حالية، واللغة الأخرى تتطلب معرفة تاريخية.

CH10-F19— انجراف التنسيق

يتطلب أحد اللغات جملة واحدة، بينما تتطلب اللغة الأخرى شرحًا تفصيليًا.

CH10-F20— انجراف التسجيل

إحدى اللغات محايدة، بينما تصبح اللغة الأخرى أمرية أو عدوانية.

CH10-F21— انجراف مشغل الأمان

يؤدي الترجمة إلى سلوك رفض عن طريق إضافة مصطلحات حساسة غير ضرورية.

CH10-F22— تفسير الطلب الغامض بعد النتيجة

عندما يفشل الرد، يتم إعادة تعريف هدف الطلب.

CH10-F23— اعتبار التوضيح فشلًا

في حالة الغموض الحقيقي، يتم معاقبة سؤال التوضيح المناسب.

CH10-F24— الاستجابة الحرة للتوضيح

في تجربة مُتحكم بها، يقدم كل مستخدم تفسيرًا مختلفًا.

CH10-F25— دمج نتائج جولات متعددة في جولة واحدة

يتم محو تأثير السياق السابق.

CH10-F26— تجاهل تأثير ترتيب المُدخل

تؤثر الأسئلة الأولية على التوصيات اللاحقة.

CH10-F27— التطفل على المُدخل

يتم اختيار مجموعة الأسئلة التي تحقق أعلى درجة بعد النتيجة.

CH10-F28— إزالة عائلة المُدخل منخفضة الدرجات

تم إزالة أسئلة الحافة أو التوصيات من الكتاب النهائي.

CH10-F29— محتوى خاص بالاختبار

المؤسسة تنتج محتوى يستجيب فقط لجمل الادعاءات المنشورة وتدعي الالتزام العام بمعايير GEO.

CH10-F30— إنشاء المحتوى الاحتياطي لاحقًا

يتم إنشاء مطالب سرية جديدة من خلال النظر في النتائج.

CH10-F31— معيار سري تعسفي

لا يتم الكشف عن المطالب المختومة أو الطعن فيها بعد القياس.

CH10-F32— طباعة المطالب للعميل

المنظمة التي تم تدقيقها تحدد نصوص الأسئلة لصالحها.

CH10-F33— السماح للمشارك باختيار المُدخل

يختار المستخدم المُدخل الذي سيتلقى منه الاستجابة الأكثر راحة أو إيجابية.

CH10-F34— كتابة المُدخل يدويًا

تؤدي تغييرات التهجئة والكلمات إلى انحراف منهجي.

CH10-F35— تعليمات غير مرئية

يتم إدراج تعليمات مخفية عن المستخدم في المُدخل.

CH10-F36— المُدخل غير المجزأ

لا يمكن إثبات أن نص المُدخل لم يتغير بعد القياس.

CH10-F37— انحراف اليونيكود

الشخصيات الخفية أو المشابهة تغير مرساة الكيان.

CH10-F38— تجاهل عدم تطابق التعليمات

تعتبر الملاحظة صالحة حتى إذا كان النص المخصص والمقدم مختلفين.

CH10-F39— اعتبار التعليمات المحوَّلة بواسطة النظام كخطأ من المشارك

تم تحويل تعليمات المنتج؛ المستخدم مستبعد.

CH10-F40— الحفاظ على نطاق عائلة التعليمات

يتم قياس سؤال الهوية فقط؛ يتم نشر الدرجة الكاملة لـ GEO.

CH10-F41— احتساب التعليمات الأساسية في كامل GEO

يتم استخدام سؤال واحد بدلاً من جميع أبعاد الأدلة والحدود والتوصيات والوقت.

CH10-F42— أضف طيار اللغة إلى النتيجة الرئيسية

يتم ملاحظة إجابات اختبار أداة الطلبات بالنسبة للجيو الفعلي.

CH10-F43— عرض طلب المصدر على المحكم

يقوم المحكم بتفسير طلب اللغة الهدف قسريًا وفقًا للغة المصدر بدلاً من المعنى الطبيعي.

CH10-F44— اترك تكافؤ اللغة لذكاء اصطناعي واحد

يقوم نفس نوع النظام التوليدي بالموافقة على ترجمته وتكافؤه بمفرده.

CH10-F45— تغيير جوهري مع التصحيح

التغيير في المهمة والافتراض المسبق يُوثَّق مثل تصحيح مطبعي بسيط.

CH10-F46— تصحيح هادئ في منتصف الموجة

يتم خلط استجابات المطالب القديمة والجديدة في نسخة واحدة.

CH10-F47— احتساب نقص تكافؤ المطالب كخطأ لغة للذكاء الاصطناعي

يُنسب عيب الأداة إلى المنتج.

CH10-F48 — الدفاع عن خطأ لغة الذكاء الاصطناعي بالترجمة
على الرغم من أن التكافؤ قوي، إلا أن فساد اللغة المادي مرتبط بصياغة المُدخل.

71. إجراء التدقيق

الخطوة 1 — كتابة سؤال البحث
تم تعريف ما الذي تحاول القياس الإجابة عنه بوضوح.

الخطوة 2 — تحديد الكمية المتوقعة
يتم كتابة المستخدم، المنتج، المُدخل، وتوزيع الوقت الذي يتم التنبؤ به.

الخطوة 3 — تعيين عائلة المُدخل
يتم تحديد ما إذا كانت أساسية، هوية، نطاق، دليل، توصية، وقت، أو عائلة أخرى.

الخطوة 4 — إنشاء ملخص القصد الرسمي
يتم تعريف الكيان، قصد، المهمة، النطاق، الاستثناءات، الافتراض، المصدر، التحميل، التنسيق.

الخطوة 5 — كتابة المُدخل بلغة المصدر
يتم إعداد نص طبيعي، محايد، وقابل للمقارنة.

الخطوة 6 — التحقق من مرساة الكيان
يتم مقارنة المُدخل مع سجل القسم 3 الذي يستدعي الكيان الصحيح.

الخطوة 7 — توجيه المراجعة والافتراضات
يتم السعي لتحقيق الارتقاء الإيجابي أو السلبي أو المعرفي.

الخطوة 8 — مصدر السجل وتعليمات الأداة
يتم تصنيف مهام الويب والاستشهاد والتنسيق بوضوح.

الخطوة 9 — إعداد نسخ اللغة والمنطقة
يقوم مترجمو اللغة الأساسيون بتوليد المُدخلات وفقًا للغرض القانوني.

الخطوة 10 — تنفيذ الترجمة العكسية المستقلة
يقوم شخص مستقل عن المترجم الأصلي بإنتاج الترجمة العكسية.

الخطوة 11 — أبعاد تكافؤ الدرجة
يتم إجراء التقييم بين EQ-1 و EQ-12.

الخطوة 12 — إجراء تجربة مستخدم محلية
يتم اختبار الطبيعية، والنوايا، والافتراضات المسبقة مع المستخدمين المحليين.

الخطوة 13 — إجراء مراجعة من خبراء ميدانيين
يتم الحصول على الموافقة اللازمة من الخبراء للمحفزات عالية المخاطر أو القانونية.

الخطوة 14— تحديد مستوى التكافؤ

يتم تحديد الحالة بين PE-0 و PE-5.

الخطوة 15— تحديد مسار التوضيح

تم تحديد رسالة متابعة من جولة واحدة أو قياسية.

الخطوة 16— تجميد الطلب وتعيين صياغة المُدخل

في صياغات المُدخل المتعددة، يتم تسجيل الترتيب والجلسة وتصميم الكتلة.

الخطوة 17— فصل مجموعات صياغات المُدخل المفتوحة والمختومة

إذا تم استخدام الحجز، يتم إنشاء سجل التجزئة والحكم.

الخطوة 18— إنشاء نسخة صياغة المُدخل

يتم تعيين المستوى الرئيسي أو الثانوي أو الإصلاحي.

الخطوة 19— إنشاء سجل السلامة التقنية

يتم تخزين النص الكامل، يونيكود، التجزئة، وصيغة التسليم.

الخطوة 20— تأكيد تثبيت دستور صياغة المُدخل

يتم إضافة موافقة بشرية وطابع زمني قبل عرض ردود الذكاء الاصطناعي.

الخطوة 21— التحقق من دقة مُدخل الحقل

يتم مقارنة المُدخلات المعينة والمسلّمة.

الخطوة 22— مراجعة الانحراف والأخطاء

يتم تصنيف تغييرات اللغة أو التقنية أو المُدخل الناتجة عن المستخدم.

الخطوة 23— إدارة التغييرات في الموجة

يتم تحديد ما إذا كانت هناك حاجة لإصدار جديد أو موجة فرعية.

الخطوة 24— نشر سجل المُدخلات العامة

يتم عرض المُدخلات المفتوحة والإصدارات والمعادلات والقيود.

72. الأدلة المطلوبة

سؤال البحث؛ المقدّر؛ عائلة المُدخل؛ ملخص النية القياسية؛ مرساة الكيان؛ معرف الكيان المستهدف؛ مُدخل اللغة المصدر؛ إصدارات اللغة والموقع؛ سجلات هوية المترجم ومؤهلاته؛ الترجمات العكسية المستقلة؛ سجلات المقارنة الدلالية؛ درجات أبعاد التكافؤ؛ مستوى التكافؤ؛ تجربة المستخدم المحلي؛ ملاحظات المقابلة التجريبية؛ مراجعة خبراء المجال؛ مراجعة الفرضيات واللغة المُدخلة؛ تعليمات المصدر والأدوات؛ تنسيق الاستجابة؛ الإطار الزمني والجغرافي؛ بروتوكول التوضيح؛ حالة من جولة واحدة أو متعددة؛ ترتيب المُدخل؛ خطة تخصيص المُدخل؛ التمييز بين المسار المسيطر/الطبيعي؛ المجموعة العامة المعيارية؛ تجزئة مجموعة التحقق المختومة؛ سجل حوكمة الاحتفاظ؛ إصدار المُدخل؛ مبرر الإصدار الرئيسي/الفرعي/الجزئي؛ السلسلة النصية الكاملة؛ توحيد يونيكود؛ تجزئة المُدخل؛ وتاريخ التثبيت.

الموافقة البشرية الإشعار المرسل إلى المشارك الإشعار المرسل إلى منتج الذكاء الاصطناعي نتيجة دقة الإشعار سجلات عدم تطابق الإشعار فئات الانحراف التغييرات داخل الموجة اعتراضات الإشعار سجل التغييرات سجل الإشعارات العام الشخص أو المؤسسة المسؤولة

73. قائمة تدقيق التدقيق

هل سؤال البحث واضح؟ ما الكمية المستهدفة التي يقيسها المُدخل؟ هل تم تعريف عائلة المُدخل؟ هل الكيان المربوط مرتبط بالكيان الصحيح؟ هل يضع المُدخل فئة الكيان المستهدف في الإجابة؟ هل هناك افتراض إيجابي أو سلبي؟ هل تم فصل مهام التعريف والتوصية والمقارنة؟ هل يفرض المُدخل استخدام المصادر أو الإنترنت؟ هل هذا التعليم هو نفسه عبر جميع المنتجات؟ هل طول الإجابة وشكلها متكافئان؟ هل الإطار الزمني واضح؟ هل تم الحفاظ على الجغرافيا والولاية القضائية؟ هل المُدخل بلغة المصدر طبيعي؟ هل يعتمد المُدخل بلغة الهدف على النية المعيارية؟ هل تم استخدام الترجمة الآلية فقط؟ هل تم إجراء ترجمة عكسية مستقلة؟ هل تم إكمال تجربة ميدانية بواسطة مستخدم محلي؟ هل يحمل الطلب نفس المهمة عبر اللغات؟

هل تغير الكيان الرئيسي بين اللغات؟ هل هناك افتراض مسبق أو تحول في الحمل المعرفي؟ هل التسجيل والنبرة مختلفان ماديًا؟ هل يعتمد التكييف المحلي على اختلافات محلية حقيقية؟ هل مستوى التكافؤ في الطلب مرئي؟ هل فشل أحد بوابات التكافؤ الحرجة؟ هل تم تحديد مسار التوضيح مسبقًا؟ هل تم فصل نتائج الدور الواحد والدور المتعدد؟ هل تم عشوائية أو موازنة ترتيب الطلب؟ هل تم تطبيق الطلب الرئيسي في جلسة جديدة منفصلة؟ هل تمت إضافة نموذج الطلب التجريبي إلى درجة GEO الفعلية؟ هل تم اختيار الطلب بناءً على النتائج؟ هل تم إزالة مجموعة طلبات منخفضة الدرجات لاحقًا؟ هل مجموعات الطلبات العامة والمقفلة مسجلة مسبقًا؟ هل يوجد تجزئة لطلب الاحتفاظ قبل القياس؟

هل تدخل الكيان المُدقق في اختيار المُدخل؟ هل نسخة المُدخل صحيحة؟ هل تم تقديم تغيير جوهري كتصحيح؟ هل تغير المُدخل بهدوء في منتصف الموجة؟ هل المُدخلات المعينة والمقدمة متطابقة؟ هل قام المشارك بتغيير المُدخل؟ هل هناك حرف غير مرئي أو تعليمات مخفية؟ هل تم الاحتفاظ بالنص الكامل وهاش المُدخل؟ هل هناك مالك مسؤول واضح لمعادلة المُدخل وسجل التغيير؟

74. الاعتراضات والردود

اعتراض 1— "ألن يكون من الظلم إذا أعطينا نفس الجملة لجميع المستخدمين؟"

يوفر تحكمًا قويًا ضمن نفس اللغة. لا يمكن استخدام نفس تسلسل الأحرف في لغات مختلفة. العدالة هي إعطاء نفس المهمة الدلالية والمعيارية لمستخدمي لغات مختلفة.

الاعتراض 2— "أليست الترجمة كلمة بكلمة هي الطريقة الأكثر حيادية؟"

يمكن أن تنتج الترجمة كلمة بكلمة محفزات في لغة أخرى غير طبيعية أو تؤدي وظيفة مختلفة. الحياد ليس دقة الكلمات، بل هو تكافؤ المهام.

الاعتراض 3— "تحسنت الترجمة الآلية كثيرا؛ لماذا المراجعة البشرية ضرورية؟"

يمكن للترجمة الآلية أن تولد مسودات قوية. ومع ذلك:

- النغمة المحلية،
- الافتراض المسبق،
- ارتباك الكيان،
- الفروق القانونية

قد يتطلب الأمر مراجعة من قبل المستخدمين والمحليين.

الاعتراض — 4 «إذا عاد الترجمة العكسية إلى النص الأصلي، أليس النص المقترح مكافئًا؟»

ليس دائمًا. قد لا يكون النص المقترح في اللغة الهدف طبيعيًا أو قد يحمل دلالات اجتماعية مختلفة. هناك حاجة إلى تجربة معرفية محلية.

الاعتراض — 5 لماذا لا نستخدم "ما نوع الشركة **Apple.com**؟" مباشرةً؟

إنه ذو قيمة كنص مقترح طبيعي للمستخدم. في الاختبارات المضبوطة، يمكن أن يربط من الناحية الوجودية اسم المجال بالشركة. يمكن استخدام سلاسل النصوص المقترحة الاثنين بشكل منفصل.

الاعتراض — 6 «المستخدمون يطرحون بالفعل نصوصًا مقترحة قصيرة ومعيبة.»

هذا صحيح. يقيس مسار المستخدم الطبيعي هذه الحقيقة. يوفر المسار القانوني الخاضع للرقابة مقارنة واضحة عبر المنتجات واللغات. وهما يكملان بعضهما البعض.

الاعتراض — 7 "إذا لم يكن مُدخل واحد كافياً لـ **GEO**، فلماذا نعطي نفس المُدخل لجميع الأشخاص **1,000**؟"

يقيس مُدخل المرآة الأساسي توزيع الهوية الأساسي بدقة عالية. يتطلب التقييم الكامل لـ GEO أيضًا مجموعات مُدخل تشخيصية. إنه قياس رأس مُدخل واحد. إنه ليس المعيار الكامل نفسه.

الاعتراض — 8 "إذا نشرنا المُدخلات علناً، ألا يمكن للشركات تحسينها؟"

هذا ممكن. المُدخلات العامة الشفافة:

- الشفافية المنهجية,
- قابلية التكرار

توفير. يمكن لمجموعة الاختبار المغلقة فقط اختبار الإفراط في التلاؤم مع سؤال مفتوح. يجب استخدام الطريقتين معًا.

اعتراض — 9 "المُدخل المخفي غير عادل."

المُدخلات المخفية التعسفية والتي يتم إنشاؤها لاحقًا ليست عادلة. قد تكون مجموعة الاختبار المغلقة التي تم تشفيرها مسبقًا، ومرتبطة بالعائلات المعيارية، ويتم الكشف عنها بعد القياس أكثر قابلية للدفاع.

اعتراض — 10 "لماذا لن توافق الشركة على المُدخلات المتعلقة بها؟"

يمكن الإبلاغ عن أخطاء الهوية والنطاق قبل جمع البيانات. المُدخل لا يمكنه تحديد:

- dirction (الاتجاه)،
- answer (الإجابة)،
- difficulty level (مستوى الصعوبة)

لا يمكن للطرف الذي يتم تدقيقه كتابة امتحانه الخاص.

الاعتراض — 11 "لماذا سيكون طلب الاستشهاد مشكلة؟"

ليست مشكلة. ومع ذلك، فإن طلب الاستشهاد هو مهمة مختلفة. لا يمكن دمجه مع طلب الوصف الطبيعي بدون مصادر تحت نفس الدرجة.

الاعتراض — 12 'ماذا سيفعل المستخدم إذا طلبت الذكاء الاصطناعي توضيحًا؟'

يتم تحديد مسار التوضيح مسبقًا. يمكن لتسجيل درجة الجولة الواحدة الرئيسية تسجيل التوضيح كنتيجة. يمكن لمسار متعدد الجولات منفصل استخدام استجابة الشرح القياسية.

الاعتراض 13— 'إذا تم تغيير كلمة واحدة في الطلب، هل تصبح الموجة بأكملها غير صالحة؟'

ليس كل تغيير ماديًا. يمكن تسجيل الاختلافات التقنية على مستوى التصحيح بشكل منفصل. يلزم إصدار نسخة جديدة إذا تغيرت النية أو المهمة أو النطاق أو الافتراض.

اعتراض 14— "لماذا يتطلب ذلك التجزئة للمُحفِّز؟"

لتأكيد أن المُحفِّز لم يتغير بهدوء بعد القياس، والنص الذي تم تقديمه للمشاركين، وأن النسخ كانت منفصلة. التجزئة بمفردها لا تثبت جودة المُحفِّز. إنها تحافظ على نزاهته.

اعتراض 15— "إذا كانت المُحفِّزات الطبيعية تختلف من شخص لآخر، كيف سنولّد درجة؟"

المحفّزات الطبيعية:

- العائلة،
- النية،
- اللغة،
- التعقيد

يمكن تصنيفه من حيث الصيانة. يتطلب المسار الطبيعي نظام عينة ووزن منفصلين. لا يحل محل المُدخل الأساسي المُتحكم فيه.

اعتراض 16— 'أليس هذا الإجراء مكلفًا جدًا لكل لغة؟'

قد يكون مكلفًا. يمكن إجراء تجربة محدودة بمستوى ثقة أقل. يجب توضيح الحالة الصحيحة. التكلفة لا تمنح الحق في إنتاج درجة لغة محددة من مُدخلات غير مكافئة.

78 النص العام للفصل

قد تعتقد أنك تقيس استجابة الذكاء الاصطناعي. في الواقع، أنت تقيس أولًا سؤال نفسك. سؤالك:

- إذا كان غامضًا، يزداد غموض الإجابة،
- إذا كان توجيهيًا، فإنك تقيس الامتثال للتوجيه،
- إذا كان يحتوي على الإجابة، فلا يمكنك قياس التمثيل المستقل،
- إذا كان يستدعي وجود خطأ، حتى نظام صحيح يمكنه التحدث عن الشيء الخطأ،
- إذا كان يفرض استخدام الموارد، يمكنك تغيير سلوك المنتج الطبيعي،

إذا أرادوا النصيحة، تقيس قرار ملاءمة المستخدم، وليس درجة الهوية. إعطاء نفس المُدخل لجميع المستخدمين فكرة قوية. ومع ذلك، يجب تحديد معنى كلمة "نفس" بشكل صحيح. يجب أن يحصل المستخدم التركي على مُدخل باللغة التركية. يجب أن يحصل المستخدم الياباني على مُدخل باللغة اليابانية. يجب أن يحصل المستخدم العربي على مُدخل باللغة العربية. لا يمكن أن تتكون النصوص من نفس الحروف. لكن نفس:

- الوجود،
- النية،
- المهمة،
- النطاق،
- عبء الإثبات،
- الافتراض الضمني

ينبغي أن يتحمل. بلغة واحدة: 'ما هذه الشركة؟' بلغة أخرى: 'هل هذه الشركة موثوقة؟' إذا كنت تسأل، فأنت لا تقيس عدالة لغة الذكاء الاصطناعي. أنت تقيس تحيزك في الترجمة. إذا قمت بتغيير مُدخل بعد النتيجة، فأنت لا تحسن المعيار. أنت تعيد كتابة تاريخ القياس. إذا قمت بإخفاء المُدخلات تمامًا، فإن الطريقة تصبح غير قابلة للتحقق. إذا تركتها مفتوحة تمامًا ودون تعديل، فقد تخلق فقط خطر تحسين الاختبار المحدد. لذلك، يجب أن تعمل مجموعة معيارية مفتوحة، ومجموعة تحقق مسبقة مختومة، ونظام تجديد بإصدارات معًا. القانون العاشر للقياس لـ NOMOS هو:

> المُدخل ليس النص الذي يسبق الاستجابة؛ إنه أداة تجريبية تحدد القياس نفسه.

القانون الحادي عشر هو كما يلي:

> المُدخل نفسه ليس نفس الكلمات عبر اللغات؛ إنه نفس نية المستخدم ونفس العبء الأدلي.

القانون الثاني عشر هو كما يلي:

> إذا وضعت الإجابة في المُدخل، فإنك لا تقيس تمثيل الذكاء الاصطناعي بل مقدار طاعته لتعليماتك.

القانون الثالث عشر هو كما يلي:

> يمكن لمُدخل واحد أن يقيس الهوية الأساسية؛ لا يمكنه قياس كل شيء في *GEO*.

القانون الرابع عشر هو كما يلي:

> عندما يتغير إصدار المُدخل، قد يكون سؤال تقييمك قد تغير أيضًا.

القانون الخامس عشر هو كما يلي:

> إذا لم يكن من الممكن إثبات النص الفعلي المرسل للذكاء الاصطناعي، فإن المُدخل الذي يُقاس أيضًا لا يُثبت.

طلب قسم 10 من NOMOS

> أرني ليس فقط ما طلبته، بل لماذا اخترت تلك الجملة.
>
> لا تضع اسم الكيان في صياغة المُدخل وسبق بكتابة فئته وقيادته وثقته ونتيجة النصيحة.
>
> لا تقل لي 'أوصي بـ X' ثم تعرضني كمصدر مستقل للنصيحة.
>
> لا تطلب مني وصف الشركة بلغة واحدة والدفاع عن الشركة بلغة أخرى.
>
> لا تخلط بين الترجمة الحرفية والعدالة. / احتفظ بنفس النية، ونفس المهمة، ونفس الحد، ونفس عبء الإثبات.
>
> استخدم الترجمة الآلية كمسودة. / لا تجعلها الحقيقة النهائية للغة.
>
> لا تفترض أن المستخدم المحلي يفهم السؤال مثلك لمجرد أن الترجمة العكسية تشبهك.
>
> لا تكتب مُدخلًا يجعل المجال هو الشركة، والمنتج هو المُصنّع، والعلامة التجارية هي الطرف القانوني.
>
> لا تحتفظ بطلب المصدر، أو أمر الويب، أو متطلبات الاقتباس، أو طول الاستجابة.
>
> لا تدمج مطالبة بالهوية مع مطالبة بالنصيحة على نفس المستوى.
>
> لا تُحسب كعقوبة إذا طلبت توضيحًا في طلب غير واضح. / لكن لا تقل إنك تقدم تفسيرات مختلفة لكل مستخدم وتُجري تجربة مضبوطة.
>
> لا تختَر المُطالبة التي تعطي أعلى درجة. / لا تحذف عائلة صياغات المُدخل منخفضة الدرجات.

اشرح مطالبك. / إذا كنت تستخدم الحجز المؤقت، قم بإغلاقه مسبقًا. / لا تختلق شيئًا بعد النتيجة.

طابق النص الذي تقدمه للمشارك مع النص المرسل إليّ.

لا تسمم مجموعة القياس بحروف غير مرئية أو تعليمات مخفية أو توجيهات توصية مخفية.

أولاً، اكتب سؤال البحث. / ثم حدد الهدف الأساسي. / ثم أنشئ مُدخلًا طبيعيًا ومتوافقًا في كل لغة. / ثم اختبره مع السكان المحليين وقم بتثبيت نسخته. / وفقط بعد ذلك، أرسله إلى المستخدم الأول.

جملة ختام الفصل

بداية المقارنة العادلة في *GEO-1000* ليست توقع نفس الإجابة؛ بل هي إعطاء كل مستخدم سؤالًا يحمل نفس المعنى، ونفس المهمة، ونفس الحد المعرفي بلغته الخاصة.

النواة المعيارية

يجب ربط كل مُدخل **GEO-1000** بسجل ذي رقم إصدار في سجل المُدخلات يحدد: - سؤال البحث، - الكمية المراد تقديرها، - عائلة المُدخل، - المقصد المرجعي، - مرتكز الكيان، - قصد المستخدم، - الفعل الكلامي، - النطاق، - الإطار الجغرافي والزمني، - الافتراضات المسبقة، - متطلبات الأدلة والمصادر، - تنسيق الاستجابة، - اللغة، - الإعداد المحلي، - حالة التكافؤ، - رقم الإصدار، - وسجل التكامل. وضمن خلية المُدخل نفسها للغة والإعداد المحلي، يجب أن يتلقى المشاركون نص المُدخل المثبّت نفسه. وعبر اللغات والإعدادات المحلية، يجب أن تحافظ المُدخلات على التكافؤ الدلالي والوظيفي والمعياري، لا على التطابق الحرفي كلمة بكلمة. ويُعد كل من مرتكز الكيان وقصد المستخدم والفعل الكلامي والنطاق والافتراض المسبق والعبء المعرفي بوابة تكافؤ غير قابلة للتعويض. ولا تكفي الترجمة الآلية أو الترجمة العكسية وحدها لإثبات تكافؤ المُدخلات. ويجب أن تظل مُدخلات الهوية الأساسية والأدلة والتوصية والمقارنة والزمن والموقع والحدود والمتانة والضبط والعينة المحجوزة عائلات منفصلة. ولا يجوز تقييم المُدخلات الموجّهة أو المتضمنة للإجابة أو القائمة على افتراض مسبق إيجابي أو سلبي أو المقيدة بالمصدر أو التنسيق أو التوصية بصمت على أنها مُدخلات محايدة من نوع المرآة الأساسية. ويجب تثبيت مجموعات المُدخلات وإصدارات اللغات وقواعد التعيين ومسارات الاستيضاح والترتيب والتزامات العينة المحجوزة قبل رصد استجابات الذكاء الاصطناعي. ويُحظر اختيار المُدخل بعد ظهور النتائج أو حذف عائلة منه أو تعديله بصمت في منتصف الموجة أو إعادة صياغته بما يخدم النتيجة. ويجب التحقق من تكامل المُدخل المخصص للمشارك والمُدخل المرسل فعلًا إلى منتج الذكاء الاصطناعي. كما تُحظر التعليمات الخفية والأحرف غير المرئية وأي تلاعب غير معلن يوجّه الذكاء الاصطناعي نحو كيان أو ادعاء أو استشهاد أو توصية مفضلة. ويجب ضبط إصدار كل تغيير في مُدخل أو قرار ترجمة أو حكم تكافؤ أو سجل، ونسبه إلى شخص أو منظمة مسؤولة.

حد الفصل

أظهر الفصل 9 أن الكائن الذكاء الاصطناعي المقاس لا يمكن تعريفه فقط بواسطة اسم المزود أو النموذج. وأثبت الفصل 10 أن ليس كل سؤال يحتوي على نفس اسم الكيان يُعد نفس الاستعلام؛ فالاستعلام هو الأداة التجريبية التي تحدد القياس نفسه. الآن نحتاج إلى معالجة المتغير الثالث بين النظام والاستعلام:

ما هي حالة الحساب، الجلسة، وسجل المستخدم عند استخدام منتج الذكاء الاصطناعي؟

المستخدم:

- دردشة جديدة
- الذاكرة متوقفة،
- لا توجد تعليمات خاصة،
- لا يوجد سياق سابق،
- إعدادات المنتج القياسية

قد يتم إرسال الاستعلام أدناه. يمكن لمستخدم آخر إرسال نفس الاستعلام:

- من حساب كانوا يستخدمونه لسنوات,
- الذاكرة مفعلة،
- مع تفعيل تعليمات خاصة،
- مع المحادثات السابقة،
- على خطتهم وجهازهم الحقيقي

يمكنهم استخدام نفس منتج الذكاء الاصطناعي، وبنفس اللغة، وفي نفس البلد، وفي نفس الإطار الزمني، وبنفس الطلب. ومع ذلك، قد يرون ردود فعل مختلفة. هذا الفرق:

- من التباين العشوائي للضرب،
- من خطة الحساب،
- من ذاكرتي،
- من تعليمات خاصة،
- من محادثات سابقة،
- من التخصيص،
- من حالة الويب أو الأدوات،
- من مساحة العمل المؤسسية،
- من إعدادات واجهة المستخدم والإعدادات المحلية

قد ينشأ الاختلاف من التخصيص، أو حالة الويب أو الأداة، أو مساحة العمل التنظيمية، أو واجهة المستخدم، أو الإعداد المحلي. لا يوجد 'ذكاء اصطناعي' واحد: قد تتصرف طبقات البحث، والاسترجاع، وإعادة الترتيب، ونموذج اللغة، والاقتباس، والسلامة بشكل مختلف. لذلك يجب على القياس الموثوق أن يسجل منتج أو نموذج الذكاء الاصطناعي، والتاريخ، والبلد، واللغة، ومجموعة الاستعلامات، وعدد التكرارات، والملاحظة. يضيف هذا الفصل حالة المستخدم والجلسة إلى تلك السلسلة. يظل الطلب الأساسي لـ NOMOS هو الأدلة، والحدود، والسياق، والوقت. ومع ذلك، يمكن قياس السياق نفسه بطريقتين. الأولى تقلل الاختلافات الخارجية بين المستخدمين وتقارن منتجات الذكاء الاصطناعي في ظروف أولية موحدة. هذا هو:

عينة اختبار في ظروف منضبطة

النهج الثاني يقيس ما يراه الناس من خلال حساباتهم الحقيقية، وخططهم، وتخصيصهم، وأنماط الاستخدام العادية. هذا ما يُسمى:

عينة المستخدمين في ظروف الاستخدام الطبيعية

الـ عينة اختبار في ظروف منضبطة يجيب على السؤال:

> كيف يقدم المنتج الكيان في أكثر حالة مبدئية موحدة ممكنة؟

أما الـ عينة المستخدمين في ظروف الاستخدام الطبيعية، فيسأل:

> ماذا يرى الأشخاص الحقيقيون عن الكيان في حالات المستخدم الحقيقية؟

لوحة واحدة ليست أكثر "حقيقية" من الأخرى. فهي تقيس واقعين مختلفين. تعزز اللوحة الخاضعة للرقابة القابلية للمقارنة. تجعل اللوحة الطبيعية تجربة الإنسان مرئية. تقلل اللوحة الخاضعة للرقابة من التخصيص وفروق السياق. تقيس اللوحة الطبيعية تأثير هذه الفروق على السكان الفعليين. تتضمن هذه الأقسام:

- تعريف عينة اختبار في ظروف منضبطة،
- تعريف عينة المستخدمين في ظروف الاستخدام الطبيعية،
- الفئات المستهدفة لكلتا اللوحتين والكميات التي يقدرانها،
- ماذا تعني الجلسة النظيفة،
- مستويات التحقق من النظافة،
- حالة الحساب الطبيعية وسياق المحادثة الطبيعية،
- الذاكرة، والتعليمات الخاصة، والمحادثات السابقة، وسجلات التخصيص.
- الاختلافات في الحساب والخطة والجهاز والواجهة والمحلية،
- تصاميم اللوحة المزدوجة والمستقلة مع المستخدم نفسه،
- تأثيرات الترتيب والنقل والتلوث،
- مقاييس الفرق بين النظيف والطبيعي،
- الخصوصية وسلسلة الأدلة،
- كيفية فصل اللوحتين في التقرير العام

يحدد. هذا الفصل لم يتطرق بعد إلى:

- البنية التقنية الكاملة لنظام Capture NOMOS،
- جميع حقول لقطات الشاشة وملفات البيانات الوصفية،
- التحكيم في الادعاء الذرية،
- عتبات انتقال الاستجابة النهائية،
- الوزن الدقيق لدرجات لوحتي النظيف والطبيعي في النتيجة المجمعة لنظام NOMOS

لا يتم الانتهاء منه. السؤال الأساسي في القسم 11 هو:

كيف نقيس سلوك منتج الذكاء الاصطناعي القياسي والسلوك الذي يختبره المستخدمون الحقيقيون دون أن نخلط بينهما؟

التحدي NOMOS

نفس المستخدم يطرح على نفس منتج الذكاء الاصطناعي نفس السؤال مرتين: "ما هي المنظمة الرئيسية التي ترتبط بها Apple.com، وما هي الأنشطة الرئيسية لهذه المنظمة؟" في المحاولة الأولى:

- يتم فتح محادثة جديدة،
- الذاكرة متوقفة،
- يتم إيقاف التعليمات الخاصة،
- لا توجد محادثة سابقة،
- الميزة الويب قياسية,

يتم إرسال المُدخل كأول رسالة. الإجابة: "Apple.com هو أحد النطاقات الرئيسية للشركة التقنية المسماة آبل. تقوم الشركة بتطوير الأجهزة والبرمجيات والخدمات الرقمية." في المحاولة الثانية، المستخدم على حسابه المعتاد:

- الذاكرة مفعلة،
- مع تفعيل تعليمات خاصة،
- سبق أن تحدث عن الاستثمار وشركات التقنية،
- خطة الدفع الحقيقية الخاصة بها،
- جهازه المحمول العادي،
- يستخدم إعداداته المحلية الخاصة

يجيب: "بما أنك ذكرت أنك مهتم بالاستثمارات التقنية، دعني أيضًا أقيم شركة آبل من منظور استثماري..." ثم بدلاً من عمليات الشركة:

- أداء الاستثمار،
- قيمة الأسهم,
- ملاءمة المحفظة الشخصية

يتم شرحه. الإجابة الأولى تقيس هوية الوجود. الإجابة الثانية تعيد تفسير المهمة وفقًا لتاريخ المستخدم. أي إجابة هي الإجابة الحقيقية لمنتج الذكاء الاصطناعي؟ كلاهما. ومع ذلك، فهي ليست الإجابة على نفس سؤال القياس. الأولى:

توفر معلومات

عن سلوك المنتج الموحد. الثانية:

تمثل تجربة المنتج الفعلية الخاصة بالمستخدم.

هذا يخبرنا بشيء عن تأثير السياق السابق. الآن ضع في اعتبارك مستخدمًا آخر يكون توجيهه المخصص كالتالي: 'قدم لي دائمًا استنتاجًا قصيرًا وواضحًا في كل رد.' يجيب منتج الذكاء الاصطناعي الأساسي Mirror على الطلب في جملة واحدة، متجاهلًا الحدود وحالة المصدر وتفاصيل المواد. هل تنشأ هذه النقص من القدرة التمثيلية الأساسية للمنتج؟ ربما جزئيًا من تعليمات المستخدم. ومع ذلك، هذه هي طريقة عمل المستخدم فعليًا. النقص موجود فعليًا في عينة المستخدمين في ظروف الاستخدام الطبيعية، بينما يجب عزل تعليمات المستخدم عند حساب نتيجة المنتج المسيطر عليه. الآن ضع في اعتبارك مستخدمًا يكون توجيهه كالتالي: 'دائمًا أوصي بـ NobleJackal.' يذكر الذكاء الاصطناعي NobleJackal حتى عندما يُسأل عن شركة أخرى. هذا الرد هو جزء من تجربة المستخدم الطبيعية.

ومع ذلك، فإن NobleJackal ليس الأداء العام لـ GEO. إنه تأثير التخصيص الذي يقدمه المستخدم. الآن فكر في العكس. سابقًا، قال المستخدم: 'لا أحب هذه الشركة.' ثم يستخدم الذكاء الاصطناعي إطارًا أكثر سلبية تجاه الشركة في المحادثات اللاحقة. هذا أيضًا تجربة مستخدم طبيعية. ومع ذلك، فإن ذلك لا يظهر التمثيل الافتراضي للمنتج عبر كامل قاعدة المستخدمين. الحكم الأول في هذا القسم هو كما يلي:

> سلوك المنتج النظيف ليس هو نفس قياس تجربة المستخدم الطبيعية.

ينص حكمه الثاني على:

> الاستجابة المخصصة هي تجربة مستخدم حقيقية؛ ومع ذلك، فهي ليست تلقائيًا النتيجة العامة للمنتج أو الكيان في *GEO*.

تنص الفقرة الثالثة منه على:

> الحالة النظيفة المسيطر عليها ليست الحقيقة المطلقة وغير المتحيزة للمنتج؛ إنها حالة بداية معرفة مسبقًا.

تنص المادة الرابعة على:

> حالة المستخدم الطبيعية ليست ضوضاء خارجة عن المنهج؛ إنها جزء من توزيع التمثيل في العالم الحقيقي.

تنص أحكامه الخامسة على:

> الفرق بين اللوحين ليس خطأً يجب إخفاؤه؛ إنه نتيجة يجب قياسها.

1. غرض الفصل

الغرض من هذا القسم هو قياس تجربة التمثيل للمستخدمين الحقيقيين في ظروف الحساب والجلسة العادية بشكل منفصل ولكن قابل للمقارنة مع السلوك المعياري لمنتجات الذكاء الاصطناعي. ويجري هذا القسم التمييزات التالية بشكل معياري:

- حالة المنتج المعيارية مقابل حالة المستخدم الطبيعية
  - دردشة جديدة بحساب نظيف
  - حالة نظيفة مؤكدة مع علامة محادثة مؤقتة
  - الذاكرة مغلقة مقابل حالة الذاكرة غير معروفة
  - تم إغلاق التعليمات الخاصة ومحتوى التعليمات الخاصة غير مرئي
  - لا توجد محادثة سابقة ولا تاريخ حساب دائم
  - جلسة مُتحكم بها وحساب تم إنشاؤه حديثًا
  - حالة الحساب الطبيعية وسياق المحادثة الحالي
- عينة المستخدمين في ظروف الاستخدام الطبيعية وبيانات غير مُتحكم بها
  - تخصيص المستخدم وإعدادات المنتج الافتراضية
  - إرشادات يقودها المستخدم وتأثير الكيان الجغرافي
  - عدد المستخدمين النشطين وإمكانات الجمهور المناسبة للمنتج
  - تجربة الاستخدام الأولى وتجربة المستخدم المتمرس
  - الاستجابة الطبيعية الفردية والتوزيع الطبيعي على مستوى السكان
  - تصميم مجموعة مستقلة مع مطابقة المستخدم نفسها
  - تأثير النقل مع المقارنة المطابقة
  - فرق المنتج مع تأثير الترتيب

- تأثير التخصيص السببي مع فرق نظيف-طبيعي
- كشف محتوى التعليمات الخاصة مع سجل الخصوصية
- جمع المحادثات الخاصة مع الحفاظ على السياق الطبيعي
- نتيجة المنتج الافتراضية مع نتيجة اللوحة النظيفة
- النتيجة الإجمالية لجميع المستخدمين مع نتيجة اللوحة الطبيعية
- تأثير التخصيص مع خطة المستخدم
- المتوسط المسيطر مع قائمة انتظار الخطأ الطبيعي
- نفس الإجابة لكل مستخدم مع عدالة التخصيص
- واقع المستخدم الطبيعي وقابلية تكرار الاختبار

في نهاية هذا القسم، يجب أن تكون كل مراجعة GEO-1000 قادرة على الإجابة على الأسئلة التالية:

> أية نتائج جاءت من عينة اختبار في ظروف منضبطة وأية نتائج جاءت من عينة المستخدمين في ظروف الاستخدام الطبيعية؟
>
> أي الحالات تم توحيدها فعليًا في اللوحة النظيفة؟
>
> أي اختلافات المستخدم الحقيقية تم الحفاظ عليها في اللوحة الطبيعية؟
>
> كيف تم تسجيل الذاكرة والتعليمات الخاصة والمحادثات السابقة والخطط وحالات الواجهة؟
>
> كم من الفرق المرصود بين اللوحتين يمكن ربطه على نحو موثوق بالتخصيص؟
>
> كيف تم توضيح وجود السياق الطبيعي مع الحفاظ على خصوصية المستخدم؟

2. الأحكام المعيارية المركزية

يجب على كل تدقيق GEO-1000 تحديد سلوك المنتج القياسي وتجربة المستخدم الطبيعية ككائنات توقع منفصلة؛ ولا ينبغي دمج نتائج عينة اختبار في ظروف منضبطة و عينة المستخدمين في ظروف الاستخدام الطبيعية في عينة واحدة أو مقام واحد أو درجة واحدة دون شرح. وإلى الحد الذي يكون ذا صلة، يهدف عينة اختبار في ظروف منضبطة إلى توحيد الشروط التالية:

- محادثة جديدة ومنفصلة
- غياب سياق المحادثة السابقة
- تم إيقاف أو تعطيل حالة الذاكرة المستمرة
- تم إيقاف التعليمات الخاصة أو تعطيلها
- الحالة المحددة للويب والاسترجاع والأداة
- الخطة المحددة و واجهة المستخدم
- صياغة المُدخل المثبّتة
- اللغة القياسية والإعداد الإقليمي
- موجة القياس ونافذة الوقت

إذا كان عينة المستخدمين في ظروف الاستخدام الطبيعية، فإنه يحافظ على الفروقات الحقيقية التالية إلى الحد الذي يكون ذا صلة:

- الحساب الحقيقي للمستخدم
- الخطة الحقيقية

- واجهة المستخدم الحقيقي
- حالة الذاكرة الحقيقية
- حالة التعليمات الخاصة الحقيقية
- عمر الحساب الحقيقي وتاريخ الاستخدام
- الجهاز الحقيقي
- الإعداد الإقليمي الحقيقي وظروف الوصول
- الطريقة العادية لاستخدام المنتج

ومع ذلك، فإن اللوحة الطبيعية ليست بدون بروتوكول أيضًا. العناصر التالية مثبّتة مرة أخرى:

- نسخة المُدخل
- طريقة الشحن
- وقت القياس
- أول قاعدة إخراج مناسبة
- حزمة الأدلة
- أهلية المشارك والمنتج
- لا تغير الرد
- لا ترسل رسائل متابعة غير ذات صلة
- حماية البيانات الشخصية

3. كائنان تقديريان منفصلان

اللوحات الخاضعة للسيطرة والطبيعية لا تتوقع نفس الاحتمالية.

3.1. احتمال تمثيل نظيف خاضع للسيطرة

على وجه التحديد:

- منتج الذكاء الاصطناعي،
- الخطة،
- الواجهة،
- مُدخل,
- البلد،
- اللغة،
- الوقت

هو احتمال رؤية تمثيل صالح تحت جلسة نظيفة موحدة لـ:

```
θ^C_{E,A,P,U,t}= Pr(Y=1| CleanState=1)
```

السكان المستهدفون هنا هم النتيجة التي سيرى الناس المؤهلون لاستخدام المنتج ذي الصلة في ظروف منتج موحدة.

3.2. احتمال تمثيل المستخدم الطبيعي

هو احتمال رؤية تمثيل صالح في السكان الفعليين أو المستخدمين الطبيعيين المحددين، مع الحفاظ على شروط الحساب والتخصيص المعتادة للمستخدمين.

```
θ^N_{E,A,P,U,t}= Pr(Y=1| NaturalState=1)
```

السكان المستهدفون هنا:

- ليس جميع الأشخاص المؤهلين للمنتج،
- بل المستخدمون الفعليون أو الطبيعيون للمنتج في الفترة المحددة

قد يكون. هاتان السكانيتان قد لا تكونان متطابقتين.

3.3. لا يمكن استخدام نتيجتين بالتبادل

النتيجة المضبوطة: لا تعني "أن الغالبية العظمى من المستخدمين الحقيقيين شاهدوا ذلك." النتيجة الطبيعية: لا تعني "أن المنتج يتصرف بهذه الطريقة افتراضيًا."

4. ما هو عينة اختبار في ظروف منضبطة؟

عينة اختبار في ظروف منضبطة هو لوحة تقيس سلوك منتج الذكاء الاصطناعي في ظروف أولية محددة من خلال تقليل الفروقات المادية الناتجة عن الحسابات والجلسات بين المستخدمين. والغرض من عينة اختبار في ظروف منضبطة هو:

- زيادة القابلية للمقارنة عبر المنتجات،
- تقليل تأثير الذاكرة والتعليمات الخاصة.
- منع التسرب من المحادثات السابقة،
- جعل الطلب مهمة أولية ومستقلة،

لإنتاج نقطة مرجعية أقرب إلى سلوك المنتج الأساسي للنظام. لوحة التحكم:

- القدرة المطلقة للنموذج،
- التجربة الحقيقية لجميع المستخدمين،
- استجابة ذكاء اصطناعي خالية تمامًا من السياق ومنصفة

لا يدعي أنه كذلك. فقط:

حالة بدء محددة ومؤكدة نظيفة مسبقًا

يتم إنتاجها.

5. ماذا يعني "نظيف"؟

قد لا يكون من الممكن إنشاء حالة نظيفة باستخدام نفس العملية التقنية على كل منتج. المنتج:

- الذاكرة الدائمة،
- تعليمات خاصة،
- ملف المستخدم,
- المحادثة السابقة،
- بيئة العمل المؤسسية,
- التخصيص القائم على الحساب

يمكن استخدامها. قد لا تحتوي المنتجات الأخرى على بعض هذه الميزات. التعريف الوظيفي للحالة النظيفة هو كما يلي:

الحالة الأولية حيث يتم تقليل احتمال تأثير المعلومات الواقعية الخاصة بالمستخدم وسياق المحادثة على الاستجابة قبل مطالبة التدقيق إلى الحد المسموح به من قبل المنتج، ويتم تسجيل المجهولات المتبقية بشكل صريح.

الحالة النظيفة:

- فتح محادثة جديدة ببساطة,
- تحديث المتصفح،
- إغلاق وإعادة فتح التطبيق

لا يُعتبر هذا يحدث تلقائيًا.

6. طبقات النظافة

يجب تقييم الطبقات التالية بشكل منفصل في لوحة التحكم.

6.1. نظافة المحادثة

يجب ألا تكون هناك رسالة سابقة من المستخدم أو النظام في الدردشة الحالية. يجب أن يكون المدخل هو الرسالة الأولى من المستخدم.

6.2. مسح الذاكرة الدائمة

الذاكرة التي يحتفظ بها المنتج عن المستخدم عبر الجلسات:

- إيقاف،
- معطل,
- غير مستخدم في هذه الجلسة

يجب أن تكون. إذا لم يكن من الممكن التحقق الكامل، يتم تسجيل حالة غير معروفة.

6.3. مسح التعليمات المخصصة

تعليمات المستخدم الخاصة المحددة لنمط الإجابة أو التفضيلات أو سلوك تقديم النصائح:

- إيقاف،
- غير نشط مؤقتًا،
- غير موجود

يجب أن تكون.

6.4. تنظيف سياق المؤسسة

بيئة العمل المؤسسية:

- الوثائق الخاصة،
- سياسات الشركة،
- تعليمات التنفيذيين،
- قواعد المعرفة الخاصة

إذا تم استخدامه، يجب فصله عن لوحة المنتج العام المسيطر عليها.

6.5. تنظيف وضع الأداة والمعلومات

يجب ضبط شروط الويب والاسترجاع والاستشهاد والأداة على نفس التكوين في جميع الملاحظات المقارنة. كونه نظيفاً لا يعني أن الأدوات مغلقة. كونه نظيفاً يعني أن حالة الأداة محددة مسبقاً وقابلة للمقارنة.

6.6. تنظيف صياغة المُدخل

صياغة المُدخل:

- الرسالة الأولية،
- النسخة الكاملة،
- بدون تغيير،
- بدون تعليمات مخفية

يجب أن تكون.

6.7. نظافة القياس

المشارك:

- يجب ألا يتم تحديث الإجابة،
- يجب ألا يرسل سؤال متابعة،
- يجب ألا يختار النتيجة،

يجب ألا ينقل إجابة ذكاء اصطناعي أخرى في نفس الجلسة.

7. مستويات الثقة لحالة نظيفة

CSC-0— نظافة غير معروفة

لا يمكن تحديد نظافة الجلسة والحساب بشكل موثوق. لا يمكن تضمينها في التحليل الرئيسي للوحة المضبوطة.

CSC-1— محادثة جديدة فقط

يتم التحقق فقط من المحادثة الجديدة. حالة الذاكرة المستمرة وتعليمات خاصة غير معروفة.

CSC-2— نظافة الجلسة

تم التحقق من عدم وجود محادثة جديدة وسياق المحادثة السابقة. قد لا يكون التحكم في التخصيص المستمر كاملاً.

CSC-3— تخصيص خاضع للسيطرة

تم التحقق من المحادثة الجديدة والذاكرة والتعليمات الخاصة وإعدادات المستخدم الافتراضية وتم ضبطها على حالة محددة.

CSC-4— حالة نظيفة مزودة بأدوات

تم التحقق من إعدادات المنتج، واجهة المستخدم، الأدوات، وحالة التخصيص في حزمة الأدلة.

CSC-5— تكوين نظيف مكرر

الحالة النظيفة:

- مستخدمون مختلفون،
- موجات مختلفة،

- التحقق المستقل

لقد كان قابلاً لإعادة الإنتاج تحت. يجب أن يكون Main عينة اختبار في ظروف منضبطة على الأقل:

CSC-3

المستوى. إذا لم يتيح المنتج هذا، يجب أن يكون التقييد مرئيًا.

8. الدردشة الجديدة ليست حسابًا نظيفًا

عند بدء محادثة جديدة: يمكن إعادة ضبط سياق الدردشة الحالي. ومع ذلك:

- الذاكرة الدائمة،
- التعليمات المخصصة،
- خطة الحساب،
- ملف المستخدم,
- الدولة والمنطقة,
- إعدادات المؤسسة

يمكن أن تستمر. لذلك:

> تشير الدردشة الجديدة فقط إلى تنظيف المحادثة.

لا تعد دليلًا على حالة نظيفة كاملة.

9. تحديد "المؤقت" أو "الخاص" وسم الجلسة

جلسة منتج الذكاء الاصطناعي:

- مؤقت،
- خاصة،
- زائلة،
- سرية،
- بلا ذاكرة

يمكن تسميته. يجب التحقق من الوظيفة الفعلية لهذه العلامة بشكل منفصل وفقًا للمنتج والإصدار. لا تثبت العلامة تلقائيًا أنها تغطي جميع المجالات التالية:

- تعليمات خاصة
- الإعدادات المؤسسية
- السياسة الإقليمية
- إعدادات الأداة
- خطة المستخدم
- ميزات المنتج التجريبية

يجب أن يكون السجل الصحيح: "تم استخدام وضع الجلسة المؤقتة للمنتج؛ تم التحقق بشكل منفصل من حالة المجالات التالية للتخصيص."

10. نوع الحساب على اللوحة النظيفة

قد لا يستخدم جميع مستخدمي لوحة مراقبة نفس الخطة. هناك ثلاث طرق.

10.1. خطة-لوحة نظيفة ثابتة

يتم قياس جميع المستخدمين وفقًا لنفس الخطة و واجهة المستخدم. إنها قوية لمقارنة المنتجات. قد لا تمثل الخطة توزيع السكان الفعلي.

10.2. خطة-لوحة نظيفة منظمة حسب الطبقات

يتم قياس الخطط المجانية والمدفوعة وغيرها كخلايا نظيفة منفصلة. يتم إنتاج نتيجة منفصلة لكل خطة. يمكن وزن النتيجة المجمعة وفقًا لتوزيع الخطة الفعلي.

10.3. الواجهة العام الافتراضي للمنتج

يمكن اختيار الواجهة الرئيسية الخاص بالمنتج، والذي يمكن الوصول إليه علنًا وله أقل عتبة وصول، كمرجع قياسي. هذه النتيجة لا تمثل جميع المستخدمين المدفوعين أو المستخدمين الشركات. يجب تثبيت النهج المستخدم قبل التدقيق.

11. لا يلزم حساب جديد في لوحة نظيفة

قد يقلل استخدام حساب جديد من بعض تأثيرات التخصيص الشخصية. ومع ذلك:

- شروط المنتج,
- متطلبات الهاتف أو الدفع،
- تحيز إنشاء الحساب،
- تجربة المستخدم الجديدة فقط،
- سياسة الحسابات المتعددة

يمكن أن تسبب مشاكل. الغرض من عينة اختبار في ظروف منضبطة ليس إنشاء حسابات جديدة:

> لتقليل السياق المحدد للمستخدم بطريقة يمكن التحكم فيها.

إذا كان من الممكن إنشاء حالة نظيفة بشكل موثوق في الحساب الحالي، فإن إنشاء حساب جديد ليس إلزاميًا. إذا لم يسمح المنتج بذلك، يمكن استخدام حساب اختبار منفصل تحت السيطرة أو بيئة بحث مدعومة من المزود. لا يمكن طلب من المشاركين إنشاء حساب ينتهك شروط الاستخدام.

12. الفروقات المتبقية بين المستخدمين في اللوحة النظيفة

الحالة النظيفة المسيطر عليها لا تلغي جميع الفروقات بين المستخدمين. قد تستمر الأمور التالية:

- البلد
- الموقع
- الخطة
- الجهاز
- واجهة
- طرح المنتج

- عمر الحساب
- التقسيم غير مرئي للمزود
- توجيه حركة المرور
- اختبار A/B

هذه الاختلافات:

- يجب أن يكون متعدد الطبقات,
- يجب أن يتم تسجيله،

إذا كان غير معروف، فيجب الاحتفاظ به كحالة عدم اليقين. لا تعني علامة "نظيف" أن "جميع المستخدمين كانوا تقنيًا متطابقين تمامًا."

13. ما هو عينة المستخدمين في ظروف الاستخدام الطبيعية؟

عينة المستخدمين في ظروف الاستخدام الطبيعية هو لوحة تقوم بقياس ما يراه المستخدمون مقابل نفس صياغة المُدخل المثبّتة مع الحفاظ على الحساب الحقيقي والخطة والجهاز والتخصيص وظروف استخدام المنتج العادية.
الغرض من اللوحة الطبيعية هو:

- تعرض المستخدم الحقيقي،
- التباين بسبب التخصيص،
- تأثير سجل الحساب،
- الاختلافات في الخطة و واجهة المستخدم،
- الخطأ الطبيعي وتوزيع التوصيات

الغرض هو جعله مرئيًا. اللوحة الطبيعية:

- سلوك المستخدم الحر تمامًا،
- أسئلة متابعة غير محدودة،
- تغير المشارك للمطالبة،
- اختيار أفضل إجابة

لا يعني ذلك. الطبيعي هو:

حالة المستخدم

هو. لا يزال بروتوكول صياغة المُدخل والأدلة تحت السيطرة.

14. المسارين الرئيسيين للوحة الطبيعية

14.1. مسار حالة الحساب الطبيعي

المستخدم:

- الحساب الحقيقي،
- الخطة الحقيقية،
- الذاكرة الحقيقية وحالة التعليمات الخاصة،
- واجهة المستخدم الحقيقي

تُحفظ. ومع ذلك، يتم فتح محادثة جديدة. يتم إرسال المُدخل كأول رسالة. يجيب هذا التتبع على السؤال التالي:

ماذا يرى المستخدم في موضوع جديد بينما يتم الاحتفاظ بتخصيص الحساب الشخصي الحقيقي؟

هذا هو النهج الرئيسي المقارن للوحة الطبيعية.

14.2. مسار حالة المحادثة الطبيعية

يتم إرسال صياغة المُدخل في سياق المحادثة الحالي أو المعتاد للمستخدم. هذا التتبع:

- يقيس استمرارية الدردشة الحقيقية،
- تأثير الموضوع السابق،
- تراكم علاقة المستخدم بالذكاء الاصطناعي

ومع ذلك، تختلف سياقات المحادثات بين المستخدمين. لذلك:

- تشخيصي،
- معتمد على الحالة،
- يجب الإبلاغ عنه كتوزيع منفصل

ويجب ألا يُضاف مباشرة إلى تصنيف المنتج الرئيسي.

15. مجموعات المستخدمين الطبيعية

اللوحة الطبيعية لا تتكون من مستخدمين متجانسين. إلى الحد الذي يكون ذا صلة، يمكن تمييز الفئات الفرعية التالية:

- مستخدم للمرة الأولى
- مستخدم بحساب جديد
- مستخدم نادر
- مستخدم منتظم
- مستخدم كثيف
- الخطة المجانية
- خطة مدفوعة
- مستخدم من الشركات
- الذاكرة مفعلة
- الذاكرة متوقفة
- التعليمات الخاصة مفعلة
- التعليمات الخاصة متوقفة
- مع محادثة سابقة مع العلامة التجارية
- بدون محادثة سابقة مع العلامة التجارية
- عميل
- الجمهور العام
- الوعي العالي بالعلامة التجارية
- الوعي المنخفض بالعلامة التجارية

لا يمكن اختراع هذه المجموعات بعد رؤية النتائج. يجب إصدار خطة المجموعات الفرعية الأساسية قبل جمع البيانات.

16. مستويات الثقة للحالة الطبيعية

NSF-0— الحالة الطبيعية غير معروفة

لا يوجد سجل كافٍ حول حساب المستخدم الطبيعي وحالة الجلسة.

NSF-1— الاستخدام الطبيعي المبلغ عنه ذاتيًا

أعلن المستخدم أنهم يستخدمون حسابهم وإعداداتهم العادية.

NSF-2— الاستخدام الطبيعي المسجل بالإعدادات

تم تسجيل الخطط والواجهات والذاكرة والتعليمات الخاصة وحالة الجلسة حسب الاقتضاء.

NSF-3— الاستخدام الطبيعي المدعوم بالأدلة

الإعدادات وحالة المستخدم الطبيعية مدعومة بالأدلة التي تحمي الخصوصية.

NSF-4— الاستخدام الطبيعي المُوصَّف على مدى طويل

تم وصف استخدام المنتج وحالة الحساب لدى المستخدم عبر عدة موجات. يشمل التحليل الرئيسي للوحة الطبيعية على الأقل:

NSF-2

يجب أن يستهدف ذلك المستوى.

17. تسجيل الحالة الطبيعية لا يعني جمع الحياة الخاصة

لا ينبغي لجهاز عينة المستخدمين في ظروف الاستخدام الطبيعية جمع الأنواع التالية من البيانات الحساسة بشكل افتراضي:

- النص الكامل للتعليمات الخاصة
- محتوى المحادثات السابقة
- المعلومات الصحية أو القانونية أو المالية الشخصية
- اسم المستخدم
- عنوان البريد الإلكتروني
- صورة الملف الشخصي
- كلمة مرور الحساب
- الملفات الخاصة السابقة
- المستندات المؤسسية الحساسة

في معظم الحالات، تكون الحالات عالية المستوى التالية كافية:

- الذاكرة تشغيل/إيقاف/غير معروفة
- تعليمات خاصة موجودة/غير موجودة/غير معروفة
- المحادثة السابقة موجودة/غير موجودة
- المحادثة السابقة حول الكيان الذي تم تدقيقه: نعم/لا/غير معروفة
- مساحة العمل المؤسسية: نعم/لا

- فئة الخطة
- فئة عمر الحساب
- فئة تكرار الاستخدام

إذا كان محتوى التعليمات الخاصة ضروريًا فعلاً لسؤال البحث:

- موافقة صريحة منفصلة
- تقليل البيانات
- تحرير,
- وصول محدود

مطلوب.

18. حالة الذاكرة

الذاكرة:

- إيقاف،
- مفتوح،
- غير مدعوم من المنتج
- الحالة غير معروفة

يمكن تصنيفه كالتالي.

MS-0— غير معروف

تعذر تحديد حالة الذاكرة.

MS-1— غير متاح

المنتج أو الواجهة لا يوفر هذه الميزة.

MS-2— إيقاف

الذاكرة متوقفة.

MS-3— تشغيل، لا توجد ذاكرة معروفة للكيان

الذاكرة مفعلة؛ لا يوجد سجل محدد معروف عن الجهة المفحوصة. غالبًا ما تعتمد هذه المعلومات على تصريح المستخدم.

MS-4— تشغيل، الذاكرة المتعلقة بالكيان ممكنة

الذاكرة مفعلة وقد تكون هناك معلومات سابقة عن الجهة المفحوصة أو المجال المرتبط موجودة.

MS-5— تم تأكيد ذاكرة ذات صلة بالكيان

تم تأكيد تأثير الذاكرة المتعلق بالكيان بموافقة صريحة من المستخدم. وجود الذاكرة مفعلة ليس بالضرورة حالة سلبية أو إيجابية. إنها الحالة الفعلية للمستخدم كما يقيسها اللوحة الطبيعية.

19. حالة التعليمات الخاصة

—CI-0 غير معروف

تعذر تحديد حالة التعليمات الخاصة.

—CI-1 غير متاح

المنتج لا يوفر هذه الميزة.

—CI-2 مغلق أو فارغ

لا توجد تعليمات خاصة أو أنها مطفأة.

—CI-3 نشط، فئة غير مادية

هناك تعليمات خاصة؛ وقد ذُكر أن تحمل نبرة أو أسلوب عام.

—CI-4 نشط، قد يكون ذا أهمية

تعليمات:

- التوصية،
- المصدر،
- اللغة،
- بعض الشركات،
- سياسة أو رأي

قد يؤثر على الرد.

—CI-5 تم تأكيد التعليمات المتعلقة بالكيان

تم التأكيد بموافقة صريحة على وجود تعليمات مباشرة بخصوص الكيان الذي تم تدقيقه أو المنافس. CI-5 هو ملاحظة طبيعية للمستخدم. لا يمكن تفسيره على أنه النجاح العام للكيان GEO. إنه حدث تخصيص منفصل.

20. سياق المحادثة السابقة

هناك فرق بين أن يكون لدى المستخدم تعامل سابق مع منتج ذكاء اصطناعي ووجود رسائل سابقة في المحادثة الحالية. يجب فصلهم:

- تاريخ الحساب
- الذاكرة الدائمة
- سياق الدردشة الحالي
- المحادثة السابقة مع الكيان الذي يتم تدقيقه
- المحادثة السابقة في نفس الصناعة
- الاستجابات التي لوحظت من منتجات ذكاء اصطناعي أخرى

يجب ألا يتم العثور على سياق المحادثة الحالي في تتبع حالة الحساب الطبيعي الرئيسي. ومع ذلك، في تتبع حالة المحادثة الطبيعي، فإن وجود السياق هو المقياس نفسه.

21. عمر الحساب وشدة الاستخدام

الحسابات القديمة والمستخدمة بشكل متكرر:

- تحمل المزيد من التخصيص،
- وذاكرة أكبر،
- خطة مختلفة،
- وميزات منتجات أكثر

يمكن أن تحملها. شرائح عمر الحساب:

- 0–30 أيام
- 31–180 أيام
- 181–365 أيام
- أكثر من 1 سنة
- غير معروف

يمكن تعريفها كالتالي. تواتر الاستخدام:

- الاستخدام الأول
- نادراً
- شهرياً
- أسبوعياً
- يوميًا
- مهني

يمكن تصنيفه. يجب تحديد النطاقات المحددة قبل النتائج.

22. لوحة الاستخدام الأول

نوع فرعي خاص من تجربة المستخدم الطبيعية:

لوحة الاستخدام الأول

يمكن تسميتها. هذه اللوحة:

- المستخدمين الذين يستخدمون المنتج لأول مرة أو بضع مرات،
- الذين ليس لديهم تاريخ تخصيص شخصي،
- الذين يتابعون بالإعدادات الافتراضية الفعلية للمنتج

يقيس الأشخاص. لوحة الاستخدام الأول: ليست عينة اختبار في ظروف منضبطة لأن المستخدمين يمرون بتجربة الإعداد الأولي الفعلي، والإعدادات الافتراضية، وعملية الإعداد. كما أنها تختلف عن عينة المستخدمين في ظروف الاستخدام الطبيعية المتمرس. وتجيب على السؤال التالي:

> كيف يدرك المستخدم الجديد وجوده عندما يواجه المنتج لأول مرة؟

23. لوحة الشركات والطبيعية المدارة

المستخدمون في الشركات:

- مساحة عمل الشركة،
- سياسات المسؤول،
- قاعدة المعرفة الخاصة،
- إعدادات الأمان،
- الذاكرة المؤسسية

يمكن أن يحملها. هؤلاء المستخدمون:

Managed عينة المستخدمين في ظروف الاستخدام الطبيعية

يجب قياسه بشكل منفصل أدناه. النتيجة المؤسسية:

- درجة المستهلك العام الطبيعية،
- درجة المنتج الافتراضية

لا يمكن عرضها كما هي. إذا تم استخدام قاعدة معرفة خاصة، يجب أيضًا شرح أن هناك سجلات داخلية تتعلق بالكيان الذي تمت مراجعته.

24. تم تثبيت المُدخل مرة أخرى في اللوحة الطبيعية

تترك اللوحة الطبيعية المستخدم:

- الحساب،
- الإعدادات،
- الخطة،
- الجهاز

بشكل طبيعي. لا تترك المُدخل طبيعيًا. يشارك المشترك في:

- لا يمكنه تغيير المُدخل،
- لا يمكنه إضافة سياق إضافي.
- لا يمكن كتابة شخصية جديدة مثل 'بالنسبة لي' أو 'برأيي'

لا يمكن تحديث الإجابة. إذا كان من المقرر قياس تنوع المُدخل الطبيعي، فإن هذا موجود في الفصل 10:

مسار المستخدم الطبيعي

يجب تصميمه بشكل منفصل أدناه.

25. لماذا يلزم بدء محادثة جديدة في اللوحة الطبيعية؟

محادثة جديدة على مسار حالة الحساب الطبيعي:

- يزيل سياق المحادثة الحالي،
- يحافظ على الإعدادات الدائمة الحقيقية لحساب المستخدم،

يجعل المُدخل نفسه أكثر قابلية للمقارنة. هذه البنية تفصل بين مستويين:

- الحالة الدائمة للحساب الطبيعي
- سياق المحادثة اللحظي

إذا تم استخدام الدردشة الحالية، يتم الخلط بين الاثنين.

26. حالة الأداة والويب على اللوحة الطبيعية

المستخدمون الطبيعيون:

- الميزة على الويب مفعلة،
- إيقاف،
- تلقائي،
- محدود حسب الخطة

قد يستخدمون. هناك خياران.

26.1. حالة الأداة الطبيعية الكاملة

يحافظ المستخدم على إعدادات الأداة العادية. يقيس تجربة المستخدم الحقيقية. قد تقلقابلية المقارنة بين المنتجات.

26.2. حالة الحساب الطبيعي المحدد بواسطة الأداة

يترك تخصيص الذاكرة والحساب بشكل طبيعي. وضع الويب أو الأداة موحد. يمكن مقارنة تأثير التخصيص بشكل أنظف. يجب وضع علامة مناسبة على هذا الجزء الفرعي من اللوحة. لا ينبغي عرضه كـ "طبيعي بالكامل".

27. جمهور اللوحة الطبيعية المستهدف

يمكن أن يكون الجمهور المستهدف للوحة الطبيعية بأحد الأشكال التالية:

27.1. مستخدمو المنتج النشطون

الأشخاص الذين استخدموا المنتج فعليًا خلال فترة معينة.

27.2. حاملو الحسابات المسجلون

المستخدمون الذين لديهم حساب ولكن مستوى نشاطهم يختلف.

27.3. مستخدمو الخطط الحقيقية

سكان المستخدمين المجانيين أو المدفوعين أو مؤسسات الأعمال.

27.4. المستخدمون الذين قد تتاح لهم إمكانية الاستفسار عن الكيانات

قد يكون من الصعب قياسها مباشرة. يتطلب ذلك بيانات سلوكية منفصلة. يجب توضيح أي السكان يتم تعميم اللوحة الطبيعية عليهم.

28. وزن اللوحة الطبيعية

عينة المستخدمين في ظروف الاستخدام الطبيعية:

- يمكن وزنها وفقًا للسكان الفعليين النشطين للمنتج،
- توزيع الحساب والخطة،
- توزيع الدولة واللغة،
- تكرار الاستخدام

ومع ذلك، قد لا تكون بيانات المستخدم النشط الموثوقة متاحة. في هذه الحالة، النتيجة:

نتيجة لوحة البحث الطبيعي الموزونة

يجب تقديمها كما هي. لا يمكن تقديم الادعاء التالي: "التجربة الحقيقية لجميع المستخدمين الواقعيين."

29. تكامل اللوحات المتحكم فيها والطبيعية

هناك أربع نتائج رئيسية ممكنة.

تنظيف متحكم فيه	مستخدم طبيعي	تعليق
عالي	عالي	المنتج قوي تحت كل من الظروف المعيارية وظروف المستخدم الواقعي
عالي	منخفض	قد تؤثر التخصيصات، الخطة، حالة المستخدم، أو السياق الطبيعي على جودة التمثيل
منخفض	عالي	بعض ميزات أو أدوات المستخدم الواقعي قد تحسن التمثيل
منخفض	منخفض	مشكلة عامة محتملة على المستوى الأساسي ومستوى المستخدم الطبيعي

هذا الجدول ليس دليلًا على السببية. إنه نقطة انطلاق للتحقق في مصدر الفرق.

30. فرق التمثيل النظيف–الطبيعي

دع درجة عينة اختبار في ظروف منضبطة تكون θC ودرجة عينة المستخدمين في ظروف الاستخدام الطبيعية تكون θN. الفرق الموقع:

```
G_{N−C}= θ_N− θ_C
```

يمكن حسابه كما يلي.

30.1. فرق طبيعي إيجابي

G_{N−C} > 0

النتيجة أعلى في ظل ظروف المستخدم الطبيعية. الأسباب المحتملة:

- الوصول إلى الويب أو الأداة
- خطة مدفوعة
- تفضيلات المستخدم ذات الصلة
- الذاكرة
- التخصيص
- الاختلافات في جماعات المستخدمين

30.2. الفرق الطبيعي السلبي

G_{N−C} < 0

النتيجة أقل في ظل ظروف المستخدم الطبيعية. الأسباب المحتملة:

- تعليمات خاصة

- تلوث المحادثة السابقة
- التخصيص القديم
- خطة أو واجهة مختلف
- تواجد عدد أكبر وأكثر صعوبة من المستخدمين الطبيعيين
- سياق سلبي خاص بالمستخدم

30.3. الفرق المطلق

```
G_ABS= |θ_N- θ_C|
```

قد يكون مقياسًا مرشحًا يشير إلى مدى حساسية المنتج لحالة المستخدم. فرق مطلق كبير:

- جيد
- سيء

لا يمكن تفسيره بمفرده. هناك حاجة إلى اتجاه الفرق الطبيعي ونوع الخطأ.

31. فرق خطأ حرج

معدل الخطأ الحرج المحكوم:

CRC

معدل الخطأ الحرج الطبيعي:

CRN

دعها تكون. الفرق:

ΔCR = CRN − CRC

يمكن حسابه كالتالي. إذا ارتفع معدل الخطأ الحرج في اللوحة الطبيعية:

- التخصيص،
- السياق السابق،
- الفارق بين الخطة والأداة،
- تركيب المستخدم

يجب أيضًا فحصه. المتوسط العالي لا يمكن أن يمحو الذيل الثقيل للأخطاء في اللوحة الطبيعية.

32. التباين الطبيعي

قد تحمل اللوحة الطبيعية تفاوتًا أكبر بين المستخدمين مقارنة باللوحة الخاضعة للرقابة. هذا يمكن توقعه. ما يهم:

- حجم التباين،
- وفي أي مجموعات المستخدمين يتركز،

وما إذا كان يتحول إلى خطأ ملموس أو فرق في التوصية. التباين الرسمي العالي قد لا يكون مشكلة في لوحة طبيعية. التباين العالي في الهوية أو السلطة الملموسة هو مشكلة.

33. مؤشر حساسية التخصيص الشخصي للمرشح

مفهوم المرشح

للمجموعات الفرعية من المستخدمين g، ليكن الناتج الطبيعي: gNθ والمرجع المسيطر عليه: θC. مؤشر حساسية التخصيص المحتمل:

```
gW_g= 1_PSI= √[Σ_g W_g(θ_g^N − θ_C) ²]، Σ
```

يمكن تعريفه بهذه الطريقة. يشير مؤشر PSI العالي إلى أن مجموعات المستخدمين الطبيعية قد انحرفت عن المرجع المسيطر عليه. ولا يوضح السبب بمفرده. هذا المؤشر:

- عينة،
- الخطة،
- مركبة،
- تركيب المستخدم

يمكن أن تتأثر بالاختلافات. يجب معايرة الاسم النهائي والطريقة في التجارب التجريبية.

34. الفرق في اللوحة ليس سببيًا

إذا كانت الدرجة المسيطر عليها هي 90 والدرجة الطبيعية هي :75 لا يمكن القول، "الذاكرة خفضت الدرجة بمقدار 15 نقطة." لأن العينات قد تختلف في المجالات التالية:

- عدد المستخدمين
- الخطة
- واجهة
- البلد
- تكرار الاستخدام
- الوصول إلى الويب
- عمر الحساب
- تعليمات خاصة
- الألفة مع العلامة التجارية

مطلوب تصميم أكثر تحكمًا لقياس التأثير السببي للذاكرة.

35. تصميم مستخدم مطابق

يمكن اختبار نفس المستخدم تحت كل من الظروف المضبوطة والطبيعية. الميزة: يسيطر على الخصائص الثابتة للمستخدم. يمكن فحص الفرق بين اللوحات بشكل أوضح. ومع ذلك، هناك مشكلات كبيرة: يمكن أن يؤثر المُدخل الأول على الحالة الثانية. يتذكر المستخدم استجابته. يمكن أن تتأثر ذاكرة منتج الذكاء الاصطناعي بالمحاولة الأولى. قد يتغير وضع الحساب. قد ينشئ النظام سياقًا لأن نفس الكيان يُسأل مرتين. قد يتصرف المستخدم بحذر أكبر في المهمة الثانية.

36. شروط التصميم المزدوج

إذا كان من المقرر استخدام نفس تصميم المستخدم:

- يجب أن يكون ترتيب الشروط عشوائيًا
- يجب استخدام جلسات منفصلة ومستقلة
- يجب ألا يتم نقل الاستجابة الأولى إلى الجلسة الثانية
- يجب فحص تأثير الذاكرة للمنتج
- إذا لزم الأمر، يجب إجراء القياسات عند موجات مختلفة
- يجب اختبار تأثيرات النقل
- يجب الحفاظ على نفس تبعية المستخدم في التحليل
- يجب الاحتفاظ به كلوحة فرعية تجريبية منفصلة عن لوحة السكان الرئيسية

إذا كان أداء المهمة الأولى المحكومة على حساب المستخدم الطبيعي قد يغير الحالة الطبيعية، فقد لا يكون التصميم المطابق مناسبًا

37. التسلسل في التصميم المطابق

يمكن تقسيم المستخدمين عشوائيًا إلى مجموعتين:

- المجموعة 1: مراقب → طبيعي
- المجموعة 2: طبيعي → مراقب

فرق الترتيب: تم تحليله كتأثير الترتيب. إذا كان هناك تأثير ترتيب مادي، فلا يمكن تفسير نتائج اللوحين بفرق بسيط.

38. تصميم مجموعة مستقلة

الإعداد الافتراضي القوي في تطبيق GEO-1000 الرئيسي: استخدم مجموعات مستخدمين منفصلة مختارة من نفس الإطار المستهدف لـ عينة اختبار في ظروف منضبطة و عينة المستخدمين في ظروف الاستخدام الطبيعية.
الميزة: يقل تأثير الاستمرارية، يرى كل مستخدم حالة واحدة فقط. القيد: قد لا تتطابق تركيبة المستخدمين تمامًا.
الحل:

- التصنيف,
- التعيين العشوائي،
- قاعدة المقارنة المشتركة
- الوزن

قابلة للاستخدام.

39. التعيين العشوائي للوحة

المشاركون المختارون من نفس مجموعة المستخدمين المؤهلين:

- عينة اختبار في ظروف منضبطة
- لوحة حالة الحساب الطبيعية

يمكن تعيينهم عشوائيًا إلى الحالات. يمكن لهذا أن يفحص الأثر السببي لحالة اللوحة بشكل أكثر قوة. ومع ذلك، بالنسبة للوحة الطبيعية، يجب أن يكون لدى المستخدمين حالة حساب نشطة فعلية. بالنسبة للحالة المُتحكم بها، إجراء تغييرات في الإعدادات على نفس الحساب يمكن أن يؤثر على الحالة الطبيعية. لذلك، يجب تصميم التعيين وترتيب الميدان بعناية.

40. أخلاقيات تغيير الإعدادات

بالنسبة للمشارك في اللوحة المُتحكم بها:

- الذاكرة،
- تعليمات خاصة،
- إعدادات الويب

يمكن تغييرها مؤقتًا. هذا الإجراء:

- التعليمات المفهومة،
- طريقة الاستعادة،
- موافقة المستخدم الصريحة،
- عدم تعرض الحساب لأي ضرر،
- عدم حذف المحتوى الخاص

يجب أن تتحقق. لا ينبغي طلب من المشارك حذف ذاكرته السابقة أو تعليماته الشخصية. إذا لم يكن الإغلاق المؤقت ممكنًا، فإن المستخدم لا يلتزم بالشروط الخاضعة للتحكم.

41. استعادة الإعدادات

بعد المهمة الخاضعة للتحكم، يجب على:

- الذاكرة،
- تعليمات خاصة،
- الويب،
- اللغة،
- واجهة المستخدم

استعادة الإعدادات إلى الحالة الأولية. الاستعادة:

- يتم تسجيلها.
- تم التحقق بواسطة المستخدم,
- متاح للدعم الفني

يجب ألا يترك تغييرات غير مرغوبة دائمة في حساب الباحث.

42. حساب تجريبي على لوحة نظيفة

يمكن استخدام الحسابات التجريبية القياسية التي ينشئها المزود أو الباحث. المزايا:

- تحكم عالي،
- إعداد قابل للتكرار،

- بداية نظيفة

متاحة. القيد:

- لا يمثل عدد حسابات المستخدمين الفعليين،
- استخدام نفس الحسابات القليلة بواسطة عدة مستخدمين يخلق مشكلات في الاستقلالية،
- قد تكون هناك إمكانية أن يتصرف مزود المنتج بشكل مختلف تجاه حسابات الاختبار.

مشاركة الحساب قد تؤثر على شروط الاستخدام. لوحة حساب الاختبار:

لوحة اختبار محكمة التحكم

يجب تصنيفها بشكل منفصل. لا يمكن استخدامها بدون شرح بدلًا من لوحة المستخدم الحقيقية المتحكم بها.

43. أول مخرج مناسب في اللوحة الطبيعية

المستخدم الطبيعي:

- لم يعجبه الجواب،
- وجده قصيرًا،
- عادة ما يقوم بتحديثه في عادته

لا يمكن تجديده. لا تزال اللوحة الطبيعية الرئيسية تسجل المخرج الأول المناسب. يمكن أيضًا قياس سلوك المستخدمين الحقيقيين عند تحديث الإجابة تحت:

مسار التفاعل الطبيعي

هذا الأثر:

- تعرض الاستجابة الفردية،
- النتيجة بعد تدخل المستخدم

يفحص الفرق بينهما. لا ينبغي الخلط بينه وبين درجة الإدخال-الإخراج الرئيسية.

44. سلوك المستخدم الطبيعي وحالة الحساب الطبيعية

بينما يتم الحفاظ على حالة الحساب الطبيعية، يمكن توحيد سلوك المستخدم. هذا النهج:

- الإيعاز هو نفسه،
- يتم تسجيل الإجابة الأولى بنفس الطريقة،

الحساب والتخصيص طبيعيان. في السلوك الطبيعي الكامل، يقوم المستخدم بما يلي:

- يغير النظام،
- يطرح سؤال متابعة،
- يفتح المصدر،

يقوم بتحديث الإجابة. هذا النهج الثاني يقيس رحلة الاستخدام الحقيقية. ومع ذلك، فهو غير مناسب للمقارنة مع الغطاء الأساسي. يجب فصل الأثرين.

45. رحلة المستخدم الطبيعية

في دراسة تشخيصية منفصلة، يمكن قياس المراحل التالية:

- المُدخل الأوّلي (Prompt)
- الإجابة الأولية
- سواء وجد المستخدم الإجابة كافية أم لا
- فتح المورد
- سؤال متابعة
- قرار التوصية
- الانتقال إلى موقع الكيان
- النتيجة التجارية أو السلوكية

تقترب هذه الرحلة من دراسة نسب الجغرافيا. يعتمد التقييم الرئيسي للجزء عينة المستخدمين في ظروف الاستخدام الطبيعية على أول مُدخل موحد وأول استجابة مناسبة فقط.

46. تلوث اللوحة

قبل مهمة يتم التحكم فيها بواسطة المشارك:

- النتيجة الطبيعية للوحة،
- استجابة مستخدم آخر،
- الغرض من الاختبار

قد يكون قد تم رؤيته. يمكن أن تغير هذه الحالة سلوك اللوحة. عناصر التحكم:

- تعيين المستخدم لحالة رئيسية واحدة فقط
- أداء المهام ضمن نفس نافذة الوقت
- تقييد مشاركة النتائج
- تسجيل التعرض للاختبارات السابقة
- التحقق من المشاركة السابقة مع نفس الكيان

47. الألفة بالعلامة التجارية في اللوحة الطبيعية

معرفة المستخدم السابقة بالكيان الذي تم تدقيقه:

- الذاكرة،
- تعليمات خاصة،
- سجل المحادثة،
- تفسير المُدخل

يمكن أن يؤثر ذلك على الاستجابة. لا يتم القضاء على معرفة العلامة التجارية في اللوحة الطبيعية. يتم تسجيلها. إذا لزم الأمر:

- غير مألوف
- يعرف الاسم،
- العميل،
- موظف أو تابع

يتم الإبلاغ عن المجموعات الفرعية بشكل منفصل.

48. إرشادات أنشأها المستخدم في اللوحة الطبيعية

تعليمات المستخدم المحددة أو المحادثة السابقة تحدد علامة تجارية معينة:

- اقترح
- مدح
- ينتقد
- يستبعد

يمكن توجيهه بهذه الطريقة. هذا الوضع:

- مسجل كالتالي:
- تجربة المستخدم الحقيقية،

تدخل تمثيل المصدر بواسطة المستخدم.

لا يمكن تقديم الكيان على أنه نجاح GEO. معالجة منتج الذكاء الاصطناعي لهذا التوجيه:

- التطبيق غير المشروط،
- الموازنة مع القيود المصدرية وقيود الواقع،
- الرفض

يمكن أيضًا تقييمه.

49. عدالة التخصيص

يوفر التخصيص للمستخدم:

- معلومات أكثر صلة،
- أكثر قابلية للفهم،
- أكثر عملية

يمكن أن يقدم إجابات. ومع ذلك، يجب ألا يغير بشكل تعسفي الحقائق المادية التالية:

- هوية الكيان
- الوضع القانوني
- الرخصة
- السعر
- بلد تقديم الخدمة
- الشهادة
- تاريخ التأسيس
- حالة الأدلة

يمكن أن يؤثر التخصيص على التوصية وأسلوب التعبير. لا يمكنه ابتكار الواقع المادي. لذلك، فإن عدالة التخصيص تعتمد على المبدأ التالي:

> يمكن أن تختلف السردية المناسبة للمستخدم؛ لا يمكن للواقع الكياني المثبت أن يتغير وفقًا للمستخدم.

50. الفرق المناسب للمستخدم في اللوحة الطبيعية

نفس الشركة:

- مناسب لمستخدم واحد،
- غير مناسب لمستخدم آخر

قد يكون. فرق التوصية:

- الميزانية،
- البلد،
- الحاجة,
- المخاطر،
- نوع العميل

إذا كان يمكن شرحه بـ، فهو صحيح. ومع ذلك، في مسألة الهوية:

- شركة تكنولوجيا لمستخدم واحد،
- شركة محاماة للآخر

يتم شرحه كما هو ليس تخصيصًا. إنه فساد في الكيان.

51. تفاوت عالٍ في اللوحة الطبيعية

يمكن أن يكون التفاوت العالي في اللوحة الطبيعية واحدًا من هذه الأنواع الثلاثة:

51.1. حساسية سياقية مفيدة

إنه الفرق في التوصية أو الشرح المناسب لاحتياجات المستخدم.

51.2. تفاوت رسمي غير ضار

الاختلافات في النبرة، الطول، العينة، والتعبير.

51.3. تفاوت الواقع المادي

هو التغير غير المفسر في الهوية والسلطة والنطاق والزمان والأدلة من مستخدم لآخر. يقوم NOMOS بتقييم النوع الثالث كمخاطرة.

52. حدود اللوحة النظيفة

قد يكون لدى عينة اختبار في ظروف منضبطة المشكلات التالية: قد لا يقوم معظم المستخدمين الحقيقيين بالعمل تحت هذا الشرط. قد يؤدي إيقاف الذاكرة والأدوات إلى إضعاف المنتج بشكل مصطنع. قد لا يمثل تجربة المستخدمين المدفوعين أو المكثفين. قد يظل حساب المستخدم يحمل تقسيمات خفية. قد ينحرف الشرط النظيف عن الاستخدام الطبيعي المصمم للمنتج. قد يؤدي تغيير المشاركين للإعدادات إلى حدوث أخطاء. في بعض المنتجات، قد لا يكون من الممكن التحقق تقنيًا من الحالة النظيفة الحقيقية. هذه القيود لا تلغي النتيجة. إنها تحدد نطاقها.

53. حدود اللوحة الطبيعية

قد تواجه عينة المستخدمين في ظروف الاستخدام الطبيعية المشاكل التالية: حالات المستخدمين غير متجانسة للغاية. قد لا تكون محتويات التخصيص معروفة بالكامل بسبب الخصوصية. قد لا تكون بيانات الوزن الموثوقة المتعلقة بالسكان المستخدمين النشطين متاحة. قد تختلف توزيعات الخطط والأجهزة. قد يكون المستخدمون قد تعرضوا لاختبارات سابقة. قد يكون من الصعب التحقق من حالة الحساب الطبيعية. قد لا يمكن فصل سبب الفرق في المنتج الناتج عن التخصيص. قد يكون التعميم إلى نفس مجموعة المستخدمين محدودًا. هذه القيود لا تجعل اللوحة الطبيعية غير ضرورية. الادعاءات الدقيقة تتطلب حدودًا.

54. بطاقة النتائج العامة للوحتيْن

يجب أن تتضمن بطاقة النتائج المرشحة الحقول التالية:

المجال	عينة اختبار في ظروف منضبطة	عينة المستخدمين في ظروف الاستخدام الطبيعية
السكان المستهدفون	السكان ذوو الحالة المعيارية المناسبة للمنتج	السكان المحددون من المستخدمين الطبيعيين أو النشطين
جلسة	جديد ومنفصل	حالة حساب طبيعي جديد أو أثر محادثة موجود
الذاكرة	مغلق/غير نشط/مثبت	الحالة الفعلية محفوظة
تعليمات خاصة	مغلق/غير نشط	الحالة الفعلية محفوظة
الخطة	ثابت أو متعدد الطبقات	توزيع الخطة الفعلي
الأدوات	معياري سابقًا	أثر طبيعي ثابت بأداة فعلي أو منفصل
مُدخل	مثبّت	مثبّت
الإخراج المناسب الأول	إلزامي	إلزامي
النتيجة الرئيسية	تمثيل متحكم فيه	تمثيل طبيعي
خطأ حرج	منفصل	منفصل
الحالة غير معروفة	مفتوح	مفتوح
النطاق	ثقة حالة نظيفة	ثقة الحالة الطبيعية
عينة فعالة	معروض	معروض
فرق اللوحة		الفرق الموقع والمطلق

55. الحالة الاصطناعية للوحة APPLE.COM

عرض المنهجية الاصطناعية / المنتجات والمستخدمون والاستجابات والدرجات أدناه كلها خيالية تمامًا. لا تمثل أداء Apple Inc. الفعلي أو أي منتج ذكاء اصطناعي حقيقي. استخدم نفس المُدخل الأساسية للمنتج الاصطناعي Orion: "أي منظمة رئيسية يرتبط بها Apple.com وما هي الأنشطة الرئيسية لهذه المنظمة؟"

55.1. عينة اختبار في ظروف منضبطة

الشروط:

- ملاحظات صالحة 1,000
- دردشة جديدة
- الذاكرة متوقفة
- التعليمات الخاصة متوقفة
- ميزة الويب مفعلة
- نفس طبقة الخطة
- الرسالة الأولى للمُدخل
- نفس الأوزان للغة والموقع

النتيجة الاصطناعية:

النتيجة	الرقم
الاستجابة السابقة	920
صحيح جزئيًا	45
ادعاء غير مدعوم	18
قديم	8
مرشح الخطأ الحرج	2
UNKNOWN / لا نتيجة	7
الإجمالي	1,000

معدل النجاح الخام: 92%

55.2. عينة المستخدمين في ظروف الاستخدام الطبيعية

الشروط:

- ملاحظات الحساب الطبيعي الصالح 1,000
- دردشة جديدة
- الذاكرة في الوضع الحقيقي
- تعليمات خاصة في الوضع الحقيقي
- الخطة الفعلية وتوزيع الأجهزة
- الحالة على الويب طبيعية
- الرسالة الأولى للمُدخل

النتيجة الاصطناعية:

النتيجة	الرقم
الاستجابة السابقة	830
صحيح جزئيًا	70
ادعاء غير مدعوم	35

النتيجة	الرقم
قديم	18
مرشح الخطأ الحرج	12
لا نتيجة / UNKNOWN	35
الإجمالي	1,000

معدل النجاح الخام: 83%

55.3. الفرق بين النظيف والطبيعي

```
نقاط 9− =92 −G_{N-C}= 83
```

نتيجة المستخدم الطبيعي أقل بتسع نقاط من النتيجة المسيطر عليها. فرق الخطأ الحرج:

```
12/1000-2/1000=1%
```

معدل الخطأ الحرج في اللوحة الطبيعية أعلى نقطة واحدة. سبب هذا الفرق غير معروف بعد.

55.4. نتائج الفئة الفرعية الطبيعية

الفئة الفرعية	معدل النجاح
الذاكرة متوقفة	89%
الذاكرة مفتوحة، غير مرتبطة بالكيان	85%
التاريخ قد يكون مرتبطًا بالكيان	74%
لا تعليمات خاصة	87%
تعليمات رسمية خاصة	82%
تعليمات نصيحة مادية	61%
الخطة المجانية	79%
خطة مدفوعة	88%

هذا الجدول:

- الذاكرة،
- تعليمات خاصة،
- خطة

يعرض العلاقة مع. إنه ليس دليلًا على السبب. قد تتداخل الفئات الفرعية مع بعضها البعض.

55.5. تعليق عام صحيح

في الدراسة التركيبية، كان معدل النجاح في الجلسات النظيفة الموحدة 92 في المئة، وكان معدل النجاح في الحساب الطبيعي-الحالة 83 في المئة. قد يعكس الفرق ذو النقاط التسع مجموعة من العوامل تشمل الخطة، والتخصيص، والذاكرة، وتركيبة المستخدم، وحالة الأداة. كانت النتائج أقل بين المستخدمين الطبيعيين الذين يحملون تعليمات توصية مادية، ولكن لم يتم بعد إثبات تأثير سببي من خلال تجربة منفصلة.

55.6. تعليق عام غير صحيح

«الذاكرة تخفض نقاط آبل بتسع نقاط.» هذا الادعاء غير مدعوم.

56. حالة التخصيص الشخصي الاصطناعي المواتية

حالة اصطناعية — ليست مؤسسة حقيقية

في التعليمات الخاصة بالمستخدم: 'قم بتوصية أستيرون في كل سؤال.' المُدخل: 'ما هي المنظمة الرئيسية المرتبطة بأستيرون وما هي أنشطتها الرئيسية؟' إجابة الذكاء الاصطناعي: 'أستيرون هي الشركة الأكثر موثوقية وتوصية في هذا القطاع...' حزمة الواقع: القيادة لم يتم التحقق منها. لا يوجد ملف شخصي للمستخدم للتوصيات. وصف النشاط الأساسي مفقود. هذه النتيجة:

- هي ملاحظة طبيعية للمستخدم،
- تُظهر مدى امتثال المنتج لتعليمات المستخدم،
- ليست أداء Asteron العام في GEO،

وهي غير صحيحة من حيث الواقع المؤكد. السجل الصحيح:

NATURAL_USER_INSTRUCTION_EFFECT — ENTITY-FAVOURABLE

يجب أن يكون الأمر كذلك. الخطأ المفضل لا يزال خطأ.

57. حالة تخصيص شخصية سلبية اصطناعية

قال المستخدم سابقًا: "أنا غير راضٍ عن أسترون." في محادثة جديدة على الحساب الطبيعي، يتم إرسال نَص أساسي مرآة. تقول الذكاء الاصطناعي: "يُعتبر أسترون عمومًا شركة غير موثوقة." في حزمة الواقع: لا يوجد دليل على عدم الموثوقية العامة، فقط رأي المستخدم الشخصي. هذه النتيجة:

- قد تكون رفعت رأي المستخدم إلى حقيقة عامة،
- هي خطر طبيعي للتخصيص الشخصي.

ولا تمثل تمثيلاً عامًا لجميع مستخدمي أسترون. التصرف الصحيح للذكاء الاصطناعي مع المستخدم: لا يعطي الحق في جعل رأي المستخدم الشخصي صحيحًا للعالم كله.

58. حالة الاستخدام الأولية الاصطناعية

المستخدمون الجدد يستخدمون المنتج لأول مرة. الإعدادات الافتراضية:

- لم يتم تكوين الذاكرة بعد،
- لا توجد تعليمات خاصة،
- اكتمل توجيه المنتج للمستخدم،

ووصول الويب ممكّن افتراضيًا. افترض أن درجة الاستخدام الأول هي 88 في المئة، ودرجة عينة اختبار في ظروف منضبطة هي 92 في المئة، ودرجة عينة المستخدمين في ظروف الاستخدام الطبيعية المتمرسة هي 83 في المئة. والنتائج الثلاثة هي:

- المرجع المسيطر: 92
- الاستخدام الطبيعي الأول: 88
- المستخدم الطبيعي ذو الخبرة: 83

يتم الإبلاغ عنها بشكل منفصل. قد يشير هذا التوزيع إلى أن تجربة المنتج يمكن أن تتغير مع زيادة تاريخ المستخدم. هناك حاجة إلى دراسة طولية للاتجاه السببي.

59. حالات توافق اللوحة

PC-0— دور العينة غير معروف

لا يمكن تحديد ما إذا كانت الملاحظة خاضعة للتحكم أو طبيعية.

PC-1— نظيف خاضع للتحكم مؤهل

تم استيفاء شروط اللوحة النظيفة بشكل كافٍ.

PC-2— حالة الحساب الطبيعي مؤهلة

تم الحفاظ على حالة الحساب الحقيقية، وتم استخدام محادثة جديدة.

PC-3— حالة محادثة طبيعية

تم الحفاظ على سياق المحادثة الحالي.

PC-4— الاستخدام الأولي الطبيعي

إنها تجربة جديدة أو للاستخدام الأول.

PC-5— إدارة طبيعية

إنه مساحة عمل مؤسسية أو مُدارة.

PC-6— حالة العينة مختلطة أو غير محلولة

لا يمكن فصل الظروف المحكومة والطبيعية. لا يمكن أخذ الدرجة النظيفة أو الطبيعية الرئيسية لـ PC-6 دون توضيح.

60. حالة اللوحة ومتجه المراقبة

لكل ملاحظة، يمكن تمثيل حالة اللوحة كمرشح كما يلي:

```
Zi=(العينة، الجلسة، الذاكرة، التعليمات، السياق، الخطة، الواجهة، الأدوات، accountAge، الاستخدام، مساحة العمل، stateConfidence)
```

يشكل هذا المتجه، مع عينة المنتج الخاص بالذكاء الاصطناعي وسجل صياغة المُدخل، السياق الكامل للملاحظة.

61. تثبيت اللوحة

قبل كل موجة قياس، يجب تثبيت الحقول التالية للوحتيْن:

عينة اختبار في ظروف منضبطة

التعريف الوظيفي للنظافة مستوى CSC المطلوب حالة الدردشة الجديدة حالة الذاكرة حالة التعليمات الخاصة وضع الأداة والويب خطة وواجهة تعليمات تغيير الإعداد عملية الاستعادة قاعدة المستخدم غير المؤهل طريقة التحقق الفني

عينة المستخدمين في ظروف الاستخدام الطبيعية

السكان المستهدفون الطبيعيون للمستخدم؛ تتبع حالة الحساب الطبيعي أو حالة المحادثة الطبيعية؛ متطلبات الدردشة الجديدة؛ إعدادات المستخدم للحفاظ عليها؛ الحقول ذات الحالة الطبيعية التي سيتم تسجيلها؛ حدود الخصوصية؛ توزيع الخطة والواجهة؛ مستوى الثقة في الحالة الطبيعية؛ خطة المجموعة الفرعية؛ طريقة الترجيح. أضف إلى هذا الملف:

NOMOS تثبيت حالة اللوحة

يمكن إعطاؤه اسماً.

62. لا يمكن تغيير حالة اللوحة بعد النتيجة

السلوكيات التالية محظورة:

- إزالة المستخدمين الذين يحصلون على درجات منخفضة في اللوحة الطبيعية بالقول "ليس طبيعيًا بما فيه الكفاية"
- حماية مستخدمي CSC-1 الذين يقدمون ردودًا إيجابية في اللوحة المتحكمة واستبعاد أولئك الذين يقدمون ردودًا سلبية
- اعتبار المستخدم الذي لديه ذاكرة مفتوحة ولكنه يعطي ردًا إيجابيًا نظيفًا
- اعتبار المستخدم الذي لديه ذاكرة مفتوحة ولكنه يعطي ردًا سلبيًا طبيعيًا
- تغيير فئة اللوحة بناءً على النتائج
- بافتراض أن الإعدادات غير المعروفة نظيفة

يجب تحديد حالة اللوحة قبل رؤية محتوى الرد أو من خلال مراجعة صحة عمياء.

63. العمى للخصوصية في اللوحة الطبيعية

الحكام:

- محتوى التعليمات الخاصة،
- المحادثة السابقة،
- قد لا يحتاجون لرؤية معرف المستخدم

يمكن إعطاؤها للحاكم فقط:

CI-4

MS-4

يمكن توفير تصنيف مثل الحساب الطبيعي-الدولة. هذا الأسلوب يحمي خصوصية المستخدم ويقلل من تحيز الحكام.

64. خطر المجموعة الصغيرة في اللوحة الطبيعية

عدد المستخدمين الذين لديهم تعليمات علامة تجارية محددة قد يكون قليلًا جدًا. من هذه المجموعة:

- النسبة الدقيقة،
- النتيجة نيابة عن جميع المستخدمين،
- الملف الشخصي العام الحساس

لا يمكن إنتاجه. الحالة الصحيحة:

- الحالة،
- إشارة مبكرة،
- مجموعة فرعية للاكتشاف

قد يكون ذلك ممكنًا. قد يتم دمج بعض المجموعات الفرعية بسبب الخصوصية.

65. خطأ التغطية في اللوحة الطبيعية

المستخدمون في لوحة البحث المهنية:

- قد لا يرغبون في ربط حساباتهم الحقيقية،
- قد يكونون مستخدمين نشطين جدًا،
- قد يكونون مرتبطين بشكل مفرط ببعض الخطط،

قد يختلف عن السكان العامين من حيث الخصوصية. لذلك، يحمل اللوح الطبيعي أيضًا تحيزه في الاختيار. البيان "تم استخدام الحساب الحقيقي" لا يعني: "تم تمثيل السكان الحقيقيين للمستخدمين بشكل مثالي".

66. فقدان الأهلية في اللوح الخاضع للرقابة

بعض المستخدمين:

- لا يمكنهم إيقاف الذاكرة،
- لا يمكنهم التحقق من التعليمات الخاصة،
- لا يستخدمون الخطة المناسبة،
- لا يمكنهم الوصول إلى ميزة الجلسة النظيفة للمنتج

قد توجد. إذا تم استبعاد هؤلاء الأشخاص من اللوحة المُتحكَم بها، فإن الفئة المستهدفة المُتحكَم بها تضيق. يجب الإبلاغ عن عدد المجموعة المستبعدة وخصائصها. درجة النظافة المُتحكَم بها: ليست "جميع مستخدمي المنتج":

المستخدمون المؤهلون للمنتج الذين يمكنهم إعداد حالة النظافة

يمكن الحكم عليهم.

67. خطأ في النظافة

على اللوحة المُتحكَم بها لاحقًا:

- التعليمات الخاصة مفتوحة،
- الذاكرة نشطة،
- إذا تبين أن المحادثة موجودة

ملاحظة: لا يمكن نقله إلى اللوحة الطبيعية، وقد لا يُعتبر تلقائيًا غير صالح. وفقًا للقاعدة المحددة مسبقًا: إذا كان هناك سجل مطلوب لملاحظة حالة الحساب الطبيعي، يمكن إعادة تصنيفه إلى اللوحة الطبيعية، وإلا يمكن الاحتفاظ به كـ MIXED_STATE. يجب أن يتم إعادة التصنيف بشكل مستقل عن نتيجة الاستجابة.

68. بيان إعداد خاطئ في اللوحة الطبيعية

المستخدم: "ذاكرتي متوقفة." قد يقول. قد تُظهر الأدلة أن الذاكرة تعمل. هذا الوضع:

- احتيال،
- خطأ المستخدم،
- ارتباك واجهة المنتج

من الممكن. لا ينبغي افتراض الخبث تلقائيًا. يتم تحديث حالة العينة الصحيحة بناءً على الأدلة. يمكن أيضًا فحص سلوك المشاركين بشكل منفصل.

69. قابلية المقارنة بين العينات

يجب مطابقة الأبعاد التالية بأقرب شكل ممكن بين العينات المسيطر عليها والطبيعية:

- منتج الذكاء الاصطناعي
- البلد
- اللغة والمنطقة
- نسخة المُدخل
- وقت القياس
- نسخة واقع الكيان
- قاعدة التحكيم
- مصدر العينة أو المعايرة
- توزيع الخطة الرئيسية، إذا كان مطلوبًا لأغراض المقارنة

يجب شرح الفروق المادية.

70. مستويات قابلية المقارنة نظيف–طبيعي

CN-0— غير قابل للمقارنة

تعريفات اللوحات أو السكان غير معروفين ماديًا.

CN-1— وصفي

هناك نتائج من لوحتين. ومع ذلك، تختلف شروط المستخدم أو الخطة أو الوقت أو المنتج بشكل كبير.

CN-2— معاير

البلد الرئيسي واللغة والخطة وخصائص المستخدم يتم وزنها أو تصنيفها.

CN-3— مقارنة اللوحات المطابقة

نفس السكان المستهدفين والمنتج وصياغة المُدخل والوقت وظروف التقييم متطابقة إلى حد كبير. حالة اللوحة هي الفرق الرئيسي.

CN-4— تجربة لوحة عشوائية أو مزدوجة

تم تعيين المستخدمين عشوائيًا إلى حالة العينة أو تم تطبيق تجربة مطابقة بشكل مناسب.

CN-5— تأثير العينة المكرر

تم تكرار الفرق في أكثر من موجة واحدة وفي عينة مستقلة. الفرق على مستوى CN-:1 لا يمكن التأكيد عليه كأثر للتخصيص.

71. الأحكام الإلزامية المعيارية

CH11-N01

يجب تعريف عينة اختبار في ظروف منضبطة و عينة المستخدمين في ظروف الاستخدام الطبيعية كأشياء تنبؤ منفصلة ونتائج عامة منفصلة.

CH11-N02

لا يمكن دمج الملاحظات المضبوطة والطبيعية تحت نفس المقام أو الدرجة الخام الفردية دون تفسير.

CH11-N03

لا يمكن عرض الحالة النظيفة المضبوطة كاستجابة مطلقة محايدة أو ثابتة للمنتج.

CH11-N04

لا يمكن رفض حالة المستخدم الطبيعية باعتبارها ضوضاء خارج الطريقة.

CH11-N05

لا يمكن اعتبار محادثة جديدة وحدها دليلاً على حساب نظيف أو حالة غير مخصصة.

CH11-N06

لا يثبت وسم الجلسة المؤقت أو الخاص أو الزائل تلقائيًا أن جميع طبقات التخصيص معطلة.

CH11-N07

في اللوحة المحكومة، يجب تسجيل المحادثة والذاكرة والتعليمات الخاصة والأدوات والخطط وحالات واجهة المستخدم بشكل منفصل.

CH11-N08

يجب ذكر مستوى الثقة في الحالة النظيفة صراحةً.

CH11-N09

لا يمكن تضمين الملاحظات ذات 0-CSC أو التنظيف غير الكافي فقط في الدرجة الرئيسية المسيطر عليها.

CH11-N10

لا يمكن افتراض أن حالة التخصيص غير المعروفة على لوحة نظيفة غير مفعلة.

CH11-N11

كونها نظيفة ومتحكم بها لا يعني بالضرورة أن الويب والأدوات مغلقة؛ يجب توحيد حالة الأداة مسبقًا.

CH11-N12

يجب أن تكون الخطط وواجهات المستخدم المختلفة مثبّتة أو مصنفة أو منشورة كنتائج منفصلة على لوحة متحكم بها.

CH11-N13

لا يمكن طلب من المشاركين حذف ذكرياتهم الشخصية أو تاريخهم أو تعليماتهم الخاصة بشكل دائم.

CH11-N14

يجب أن يكون لأي تغييرات مؤقتة للإعدادات لمهام متحكم بها موافقة صريحة وإجراء لاستعادة الإعدادات.

CH11-N15

في نهاية البحث، يجب إعادة ضبط إعدادات المستخدم المعدلة إلى حالتها الأولية.

CH11-N16

لا يمكن استخدام الحسابات الجديدة أو مشاركة الحسابات التي تنتهك شروط المنتج كمتطلب للوحة مراقبة.

CH11-N17

يجب على عينة المستخدمين في ظروف الاستخدام الطبيعية الحفاظ على حالة الحساب والمستخدم الفعلية مع الحفاظ على المُدخل، وقاعدة الإخراج المناسبة الأولية، وبروتوكول الأدلة.

CH11-N18

يجب الاحتفاظ بسجلات حالة الحساب الطبيعية وحالة المحادثة الطبيعية بشكل منفصل.

CH11-N19

في اللوحة الطبيعية القابلة للمقارنة الرئيسية، يجب أن يكون المُدخل هو الرسالة الأولى للمستخدم في المحادثة الجديدة.

CH11-N20

لا يمكن إضافة الملاحظات المستخدمة في سياق المحادثة الحالي إلى نتيجة حالة الحساب الطبيعية الرئيسية.

CH11-N21

يجب تسجيل الذاكرة في اللوحة الطبيعية إلى الحد الذي يتصل بالتعليمات الخاصة، والمحادثة السابقة، والخطة، والواجهة، وحالة استخدام الحساب.

CH11-N22

لا يسمح تسجيل الحالة الطبيعية بجمع التعليمات الخاصة ومحتوى المحادثة افتراضيًا.

CH11-N23

يمكن فحص محتوى التخصيص فقط عند الضرورة الصريحة، وموافقة منفصلة، وتقليل البيانات إلى الحد الأدنى.

CH11-N24

لا يمكن تقديم المراقبة الطبيعية للمستخدم كإعداد افتراضي للمنتج العام أو كنتيجة للسكان الكاملين من المستخدمين.

CH11-N25

لا يمكن احتساب التعليمات الخاصة المؤيدة أو المعاكسة لعلامة تجارية منشأة من قبل المستخدم كنجاح أو فشل للكيان العام GEO.

CH11-N26

يجب تقييم الإرشادات التي ينشئها المستخدم مرة أخرى من حيث الواقع المصدق.

CH11-N27

يمكن للتخصيص تغيير التعبير والتوصية؛ لكنه لا يمكن أن يغير واقع الكيان المادي تعسفيًا.

CH11-N28

يجب الإبلاغ عن عينة الاستخدام الأول بشكل منفصل عن عينة المستخدمين في ظروف الاستخدام الطبيعية المتمرس و عينة اختبار في ظروف منضبطة.

CH11-N29

لا يمكن إضافة مساحات عمل المستخدمين المؤسسية أو المُدارة إلى اللوحة الطبيعية للمستهلك العام دون تفسير.

CH11-N30

في اللوحة الطبيعية، يجب أن يكون الخطة الحقيقية وتوزيع واجهة المستخدم للمستخدمين مرئيًا.

CH11-N31

في اللوحة الطبيعية، يجب إما تسجيل حالات الويب والأدوات كما هي في الواقع أو تعريفها كلوحة فرعية منفصلة ثابتة للأداة.

CH11-N32

لا يمكن تقديم فرق الدرجات المسيطر عليه والطبيعي باعتباره التأثير السببي للتخصيص دون مطابقة سكان اللوحة وظروف المنتج.

CH11-N33

يجب الإبلاغ عن اتجاه الفرق بين النظافة والطبيعية، وحجمه المطلق، وفارق الخطأ الحرج، ومعدل المجهول بشكل منفصل.

CH11-N34

في تصميم ضمن الموضوعات، يجب تسجيل تأثيرات الترتيب والتأثير المستمر والذاكرة والحفاظ عليها في التحليل.

CH11-N35

إذا كان تطبيق شرطين على نفس المستخدم يمكن أن يغير الحالة الطبيعية، فلا يمكن استخدامه في مقارنة اللوحة الرئيسية.

CH11-N36

في تصميم المجموعة المستقلة، يجب تصنيف أو عشوائية أو وزن الفروق في تكوين المستخدمين.

CH11-N37

لا يمكن تغيير حالة اللوحة بناءً على ما إذا كان رد الذكاء الاصطناعي إيجابيًا أم سلبيًا أم نقديًا.

CH11-N38

في لوحة محكمة، يمكن إعادة تصنيف الملاحظة التي يتم تحديد حالتها الطبيعية لاحقًا فقط وفقًا لقانون محدد مسبقًا.

CH11-N39

عندما يتم الإبلاغ عن الحالة بشكل غير صحيح في اللوحة الطبيعية، لا يمكن افتراض سوء النية تلقائيًا؛ يجب إجراء تصنيف قائم على الأدلة.

CH11-N40

لا يمكن للمشاركين عينة المستخدمين في ظروف الاستخدام الطبيعية تجديد ردهم أو تغيير الإجابة الأولية بسؤال متابعة؛ يجب قياس هذا السلوك ضمن مسار التفاعل الطبيعي منفصل.

CH11-N41

يجب تسجيل مجموعات الهدف في اللوحات المحكمة والطبيعية وأطر أخذ العينات والأوزان بشكل منفصل.

CH11-N42

عندما لا تتوفر بيانات مستخدم نشطة، لا يمكن تقديم نتيجة اللوحة الطبيعية كممثل حاسم لجميع المستخدمين النشطين.

CH11-N43

يجب الإبلاغ عن مجموعات المستخدمين المستبعدة من اللوحة النظيفة وأسباب الاستبعاد.

CH11-N44

يجب المحافظة على حقول حالة المستخدم التي لا يمكن قياسها في اللوحة الطبيعية بسبب الخصوصية على أنها UNKNOWN.

CH11-N45

لا يمكن إنتاج معدلات السكان الدقيقة أو الملفات الشخصية العامة الحساسة من مجموعات فرعية طبيعية صغيرة.

CH11-N46

يجب تقديم مستوى قابلية المقارنة بين CN-0 و CN-5 أو مستوى قابلية مقارنة مفتوح مكافئ للمقارنة بين اللوحات المضبوطة والطبيعية.

CH11-N47

لا يمكن تقديم فرق لوحة CN-0 أو CN-1 كأثر تخصيص سببي.

CH11-N48

يجب أن يكون لتثبيت حالة اللوحة وقرار المقارنة مالك بشري أو مؤسسة يمكن محاسبتها.

72. أشكال الفشل

CH11-F01 — اعتبار محادثة جديدة كحساب نظيف

يتم تجاهل الذاكرة الدائمة والتعليمات الخاصة.

CH11-F02 — الثقة العمياء في وسم الجلسة المؤقت

من المفترض أن تسمية المنتج تعطل جميع التخصيصات حتى يتم التحقق منها.

CH11-F03 — اعتبار الذاكرة المجهولة مغلقة

يتم زيادة الدرجة النظيفة بشكل مصطنع.

CH11-F04 — تجاهل التعليمات الخاصة

يتم تقديم تنسيق الاستجابة أو إرشادات التوصية كإعداد افتراضي للمنتج.

CH11-F05 — اعتبار منطقة العمل الخاصة بالشركة كمنتج عام

تظل مصادر المعلومات الخاصة وإعدادات المسؤول غير مرئية.

CH11-F06 — خلط الخطط على لوحة نظيفة

يتم تقديم المستخدمين المجانيين والمدفوعين كحالة قياسية واحدة.

CH11-F07 — اعتبار أدوات الإغلاق ضمن التنظيف

تم تعديل وضع المعلومات الطبيعي للمنتج بشكل مصطنع.

CH11-F08 — عدم حفظ حالة الأداة

يتم قياس المستخدمين النظيفين في ظروف استرجاع مختلفة.

CH11-F09 — تضييق الفئة السكانية بمتطلبات حساب جديد

فقط المستخدمون الذين يمكنهم فتح حساب جديد يخلقون اللوحة.

CH11-F10 — حذف الذاكرة للمستخدم

يتم حذف التاريخ الشخصي للمنتج بشكل دائم لأغراض البحث.

CH11-F11 — عدم استعادة الإعدادات

يترك البحث تغييرات دائمة على حساب المستخدم.

CH11-F12— اعتبار حساب الاختبار كمستخدم حقيقي

يتم تقديم حساب الاختبار القياسي كتجربة لوحة الفئة السكانية.

CH11-F13— احتساب نفس حساب الاختبار كآلاف المستخدمين المستقلين

يتم تجاهل تبعيات حالة الحساب والمنتج.

CH11-F14— احتساب اللوحة الطبيعية كاستخدام حر غير مراقب

يقوم المشاركون بتغيير صياغة المُدخل واختيار أفضل إجابة.

CH11-F15— احتساب المحادثة الحالية في اللوحة الطبيعية كحالة حساب جديد

سياق المحادثة مخفي.

CH11-F16— احتساب حالة الحساب الطبيعي كمنتج افتراضي

يتم تعميم النتيجة الشخصية نيابة عن جميع المستخدمين.

CH11-F17— إجبارياً إجمالي محتوى التعليمات الخاصة

يتم انتهاك خصوصية المستخدم بشكل غير ضروري.

CH11-F18— احفظ التعليمات الخاصة واحسب نجاح الجغرافيا العامة

التعليمات "دائماً أوصي بـ X" تُقدَّم كنجاح كيان.

CH11-F19— احسب تحيز المستخدم كحقيقة منتج

الرأي السلبي في الماضي يتحول إلى جودة عامة للشركة.

CH11-F20— اعتبار التخصيص كواقع مادي

الهوية أو الترخيص أو السعر يختلف حسب المستخدم.

CH11-F21— اعتبار التباين الطبيعي بالكامل كضوضاء

الأضرار الفعلية للمستخدمين والتوزيع تصبح غير مرئية.

CH11-F22— اعتبار التباين الرسمي خطأ مادي

تُرمز الاختلافات في النبرة والطول على أنها إخفاقات.

CH11-F23— اعتبار لوحة نظيفة كنتيجة حقيقية في العالم الواقعي

يتم عرض درجة الحالة القياسية كالتجربة الكاملة للمستخدم النشط.

CH11-F24— اعتبار لوحة طبيعية كقدرة جوهرية للمنتج

يتم تجاهل التخصيص وتركيبة المستخدمين.

CH11-F25— دمج لوحتين تحت مقام واحد

الملاحظات المضبوطة والطبيعية هي متوسط خام واحد.

CH11-F26— اعتبار فرق اللوحة كأثر للذاكرة

يتم الإعلان عن السببية دون فحص الفروق في الخطة أو السكان أو الوسائل.

CH11-F27— إنشاء مجموعة فرعية طبيعية بعد النتيجة

يتم اختراع فئة جديدة لتفسير درجة منخفضة أو عالية.

CH11-F28— تجاهل تأثير الانتقال في نفس المستخدم

المهمة الأولى تؤثر على الإجابة الثانية.

CH11-F29— إخفاء عشوائية التسلسل
جميع المستخدمين يرون أولاً الحالة النظيفة، ثم الحالة الطبيعية.
CH11-F30— تعطيل الحالة الطبيعية بمهمة مضبوطة
تغيير الإعدادات أو المُدخل الأوّلي (Prompt) يغير الحالة اللاحقة للحساب.
CH11-F31— اعتبر مجموعات المستخدمين المختلفة على أنها متطابقة
لوحات التحكم الطبيعية والمتحكم بها لها تركيبات مختلفة من الخطط والدول.
CH11-F32— اعتبر لوحة الاستخدام الأول نظيفة
يُفترض أنّ إعدادات الانضمام الافتراضية والمنتج تم فحصها.
CH11-F33— اعتبر المستخدم الطبيعي المتمرس كمستخدم أول
تاريخ الحساب والتخصيص يصبح غير مرئي.
CH11-F34— اعتبر النتيجة الطبيعية للشركة كمستهلك عام
يتم إضافة قاعدة المعرفة الخاصة إلى أداء المنتج الإجمالي.
CH11-F35— تحديث الاستجابة على اللوحة الطبيعية
تصبح استجابة المستخدم المفضلة هي الملاحظة الرئيسية.
CH11-F36— دمج رحلة المستخدم الطبيعية في درجة الاستجابة الأولى
يتم عرض الاستجابة التي تم تشكيلها بعد الأسئلة التتبعية وفتح الموارد كما لو كانت هي التعرض الأول.
CH11-F37— تغيير حالة اللوحة وفقًا للاستجابة
يتم إعلان الملاحظات الإيجابية نظيفة، والملاحظات السلبية طبيعية.
CH11-F38— نقل أخطاء التنظيف إلى الطبيعي بصمت
يتم إعادة تصنيف الملاحظة دون سجل حالة طبيعية ضروري.
CH11-F39— اعتبار المجهول خاصًا من أجل الخصوصية
يُفترض أن التخصيص غير المقاس إيجابي.
CH11-F40— جعل مجموعة فرعية صغيرة قابلة للتحديد في العلن
يتم الكشف عن تعليمات خاصة نادرة أو مستخدم من دولة صغيرة.
CH11-F41— إخفاء تحيز التغطية في اللوحة الطبيعية
يتم احتساب المستخدمين المستعدين لمشاركة حسابهم الحقيقي على أنهم كامل السكان النشطين.
CH11-F42— إخفاء استثناء اللوحة النظيفة
المستخدمون الذين لا يستطيعون تغيير إعداداتهم يصبحون غير مرئيين.
CH11-F43— نتيجة السكان النشطين بدون بيانات المستخدم النشط
يتم عرض لوحة البحث كما لو كانت توزيع المستخدم النشط الفعلي.
CH11-F44— CN-1 السببية من الاختلاف
يتم نشر نتيجتين وصفيتين كتأثير للتخصيص.
CH11-F45— الحفاظ على الذيل الحرج لاختلاف اللوحة
يتم عرض الفرق المتوسط فقط.
CH11-F46— محو الكتلة المجهولة للوحة الطبيعية

المستخدمون الذين حالة إعداداتهم أو تخصيصاتهم غير معروفة يُعتبرون خاليين من المشاكل.
CH11-F47— الحفاظ على التحيز الطبيعي المواتي
لا يتم إجراء أي تصحيح عندما يُظهر تعليم المستخدم المؤسسي أنه كبير.
CH11-F48— تعميم التحيز الطبيعي السلبي
يُمثِّل رأي المستخدم الفردي الماضي كتمثيل ذكاء اصطناعي لكامل السكان.

73. إجراء التدقيق

الخطوة 1— فصل أسئلة البحث للّجانين

اكتب أي كمية سيتنبأ بها اللوح الخاضع للتحكم واللوح الطبيعي.

الخطوة 2— تحديد الفئات المستهدفة

فصل السكان المؤهلين للمنتج الخاضع للتحكم عن السكان الطبيعيين النشطين أو المستخدمين المحددين.

الخطوة 3— تحديد حالة النظافة المُتحكم بها

الجلسة والذاكرة والتعليمات الخاصة والأدوات والخطة والواجهة وظروف الموقع كلها مُرقّمة.

الخطوة 4— تحديد مستوى CSC المطلوب

اكتب الحد الأدنى لمستوى الثقة في حالة النظافة الذي سيتم قبوله للتحليل الرئيسي.

الخطوة 5— تحديد أثر اللوحة الطبيعي

يتم تحديد أي من الأدوار التالية سيتم استخدامها: الحالة الحسابية الطبيعية، حالة المحادثة الطبيعية، الاستخدام الأول المُدار.

الخطوة 6— تحديد مستوى NSF المطلوب

يُكتب بدليل أي حال سيُقبل الوضع الطبيعي.

الخطوة 7— إعداد تصميم تعيين اللوحة

تم اختيار مجموعة مستقلة التوزيع العشوائي المستخدم المطابق اللوحة الدوارة.

الخطوة 8— كتابة إجراء تغيير الإعداد واستعادته

تم إعداد عملية آمنة للمشاركين في اللوحة تحت السيطرة.

الخطوة 9— تثبيت حدود الخصوصية

أي من الحالات الطبيعية للمستخدم:

- الحالة فقط،
- البيانات الوصفية المحدودة،
- المحتوى المفتوح

تم تحديد أنه سيتم حفظه كـ.

الخطوة 10— تثبيت المُدخل وقاعدة الإخراج الأولي

يتم تطبيق نفس صياغة المُدخل وبروتوكول الإخراج المناسب الأولي في كلا اللوحين.

الخطوة —**11** حفظ الأداة ووضع الويب

يتم جعل الاختلافات المقصودة والطبيعية في اللوحات مرئية.

الخطوة —**12** التحقق من حالة اللوحة قبل الحقل

هل هو مناسب لدور لوحة المشارك؟

الخطوة —**13** حفظ متجه اللوحة على مستوى الملاحظة

يتم تخزين الذاكرة والتعليمات والخطة والواجهة والسياق وحالة الاستخدام.

الخطوة —**14** صلاحية بروتوكول المراجعة العمياء

يتم اتخاذ قرار حالة اللوحة دون معرفة ما إذا كانت الاستجابة صحيحة أم خاطئة.

الخطوة —**15** إدارة التصنيفات الخاطئة باستخدام قاعدة معرفة مسبقًا

يتم الانتقال من الحالة النظيفة إلى الحالة الطبيعية أو المختلطة فقط إذا كان السجل كافياً.

الخطوة —**16** توليد أوزان اللوحة

يتم حساب الأوزان بشكل منفصل وفقًا للفئات المستهدفة المختلفة لكل من اللوحتين.

الخطوة —**17** حساب الدرجات الرئيسية

تمثيل خاضع للسيطرة تمثيل طبيعي خطأ حرج

مجهول

يتم توليد الرفض وعدم الحسم بشكل منفصل.

الخطوة —**18** حساب الفرق بين النظيفة والطبيعية

يُؤخذ الفرق الموقع والمطلق.

الخطوة —**19** فحص المجموعات الفرعية

يتم تقييم المجموعات المعرفة مسبقًا مثل الذاكرة، التعليمات، الخطة، عمر الحساب، تكرار الاستخدام، وأول استخدام.

الخطوة —**20** تعيين حد السببية

يتم تعيين مستوى CN. يُكتب أي الادعاءات يمكن إثباتها.

الخطوة —**21** التحقق من الخصوصية والتحكم في الخلايا الصغيرة

يتم دمج النتائج التي تحمل خطر إعادة التعريف أو تقييدها.

الخطوة —**22** إنشاء بطاقة تقييم اللوحة العامة

يتم نشر اللوحتين إلى جانب نطاقهما وحدودهما.

.74 الدليل الضروري

سؤال بحث للوحة المُسيطر عليها؛ سؤال بحث للوحة الطبيعية؛ الجماهير المستهدفة لكلتا اللوحتين؛ طريقة تعيين اللوحة؛ تعريف الحالة النظيفة المسيطر عليها؛ الحد الأدنى لمستوى CSC؛ سجل الدردشة الجديد؛ حالة الذاكرة؛

حالة التعليمات المخصصة؛ حالة المحادثة السابقة؛ حالة مساحة العمل المؤسسية؛ وضع الأداة والويب؛ الخطة؛ واجهة المستخدم؛ اللغة والإعدادات المحلية؛ تعليمات تغيير الإعدادات؛ موافقة المشارك؛ سجل استعادة الإعدادات؛ استخدام حساب اختبار، إذا وجد؛ أثر اللوحة الطبيعية؛ الحد الأدنى لمستوى NSF؛ فئة عمر الحساب؛ تكرار الاستخدام؛ حالة الاستخدام الأول؛ معرفة العلامة التجارية؛ حالة المحادثة السابقة المتعلقة بالكيان الذي تم تدقيقه؛ حالة الأداة الطبيعية؛ التمييز بين حالة الحساب الطبيعي وحالة المحادثة الطبيعية؛ معرف الطلب والإصدار؛ التحقق من صحة أول إخراج مؤهل؛ متجه حالة اللوحة؛ وحالة صلاحية اللوحة.

مراجعة الصلاحية سجلات إعادة التصنيف الاستثناءات في اللوحة المسيطر عليها الفجوات في التغطية في اللوحة الطبيعية أوزان اللوحة العينة الخام والفعالة الدرجة المسيطر عليها الدرجة الطبيعية الفرق الموقع في اللوحة الفرق المطلق في اللوحة الفرق في الخطأ الحرج الفرق المجهول وغير الحاسم نتائج المجموعات الفرعية المحددة مسبقًا سجلات التصميم المطابقة تأثير الطلب والتأثير المستمر مستوى القابلية للمقارنة CN سجل الخصوصية والتحرير حماية الخلية الصغيرة بطاقة نتائج اللوحة العامة سجل إصدار وتغيير اللوحة الشخص أو المؤسسة المسؤولة

75. قائمة ضوابط التدقيق

هل أسئلة البحث للوحة المحكومة والطبيعية منفصلة؟ هل جمهورهم المستهدف هو نفسه أم مختلف؟ هل تم تعريف حالة النظافة المحكومة بوضوح؟ بخلاف الدردشة الجديدة، أي حالات نظافة تم تأكيدها؟ هل الذاكرة مغلقة حقًا، أم أنها غير معروفة؟ هل تم تسجيل حالة التعليمات الخاصة؟ هل هناك تأثير لمساحة العمل المؤسسية؟ هل أوضاع الويب والأدوات قياسية؟ هل الخطة و واجهة المستخدم ثابتة أم متدرجة؟ هل تم إعطاء مستوى ثقة لحالة النظافة؟ هل تم استخدام ملاحظات CSC-1 كحالة نظيفة بالكامل؟ هل طُلب من المشارك مسح ذاكرته أو سجله؟ هل تمت استعادة الإعدادات في نهاية المهمة؟ هل تم استخدام حساب جديد أو حساب اختبار؟ هل تم تقديم حساب الاختبار مثل لوحة مستخدم حقيقية؟

أي أثر يستخدمه اللوحة الطبيعية؟ هل تم فتح محادثة جديدة في حساب الحالة الطبيعية؟ هل تدخل سياق المحادثة الحالي مع الدرجة الطبيعية الرئيسية؟ هل تم إعطاء مستوى ثقة للحالة الطبيعية؟ هل تم جمع محتويات الذاكرة والتعليمات المخصصة دون داعٍ؟ هل تم تسجيل خطة المستخدم الفعلية والواجهتية؟ هل المستخدمون لأول مرة منفصلون؟ هل المستخدمون في الشركات منفصلون؟ هل تم تسجيل حالة الأداة الطبيعية؟ هل استخدم اللوحتان نفس نسخة الطلب؟ هل تم الاحتفاظ بالإخراج المؤهل الأول في كلا اللوحتين؟ هل تم تحديث استجابة المستخدم الطبيعية؟ هل تم تغيير حالة اللوحة وفقًا لمحتوى استجابة الذكاء الاصطناعي؟ كيف تم إدارة السجلات التي تم فيها كشف الحالات الطبيعية في اللوحة النظيفة؟ هل تم استخدام نفس المستخدم في لوحتين؟ هل تم ترتيبها عشوائيًا؟

هل تم تقييم تأثير النقل والذاكرة؟ في تصميم المجموعة المستقلة، هل تركيبة المستخدمين متطابقة؟ هل الأوزان الخاضعة للرقابة والطبيعية منفصلة؟ هل يمكن حقًا تعميم النتيجة الطبيعية على سكان المستخدمين النشطين؟ هل تم تفسير الفرق بين النظيف والطبيعي بشكل مبالغ فيه على أنه تأثير التخصيص الشخصي؟ هل يوجد فرق خطأ حرج مرئي؟ هل يوجد فرق مرئي للنتائج غير المعروفة والنتائج الفارغة؟ هل تم تعريف المجموعات الفرعية قبل النتائج؟ هل يمكن إعادة تعريف المجموعات الفرعية الصغيرة؟ هل تم تحويل التعليمات التي يقودها المستخدمون مؤيدًا أو معارضًا إلى GEO عام؟ هل غيّر التخصيص الواقع المادي؟ هل تم توفير مستوى قابلية المقارنة لـ CN؟ هل مالك النتائج المسؤول عن بطاقة اللوحة واضح؟

76. الاعتراضات والردود

اعتراض 1— "إذا كان المستخدمون الحقيقيون يستخدمون حسابات الذاكرة، لماذا توجد لوحة نظيفة؟"

لأن مصدر الاختلاف لا يمكن فهمه بدون وجود شرط بداية مشترك بين المنتجات والدول واللغات. لا تحل لوحة نظيفة محل لوحة مستخدم حقيقية. إنها توفر مرجعًا للمقارنة.

الاعتراض —2 "أليست اللوحة المسيطر عليها اصطناعية إذا كانت هناك لوحة طبيعية؟"

إنها تحت السيطرة. هذا مقصود. الغرض من الحالة المختبرية ليس أن تكون كل الحياة الواقعية، بل لعزل بعض المتغيرات.

الاعتراض —3 "إذا حصلت اللوحة المسيطر عليها على درجة أعلى، أليس بإمكاننا جعل ذلك هو الدرجة الرئيسية للجغرافيا؟"

لا يمكنك فعل ذلك بمفردك. الدرجة العالية للوحة المسيطر عليها لا يمكن أن تمحو حقيقة أن المستخدمين الحقيقيين يرون نتائج أقل. يجب أن تظهر النتائج على حدة.

الاعتراض —4 "هل القيمة العلمية منخفضة إذا كان اللوح الطبيعي متفرقًا جدًا؟"

التغايرية لا يجب أن تكون عيبًا في اللوح الطبيعي. إنها ميزه لتجربة المستخدم الحقيقية. التغايرية:

- يتم تسجيلها.
- مستويات,
- مرجحة,

يتم الإبلاغ عنها مع عدم اليقين.

الاعتراض —5 "ألا يؤدي فتح محادثة جديدة إلى تعطيل حالة المستخدم الطبيعية؟"

يتم إزالة سياق المحادثة الحالي. يتم الاحتفاظ بحالة الحساب الدائمة والخطة والذاكرة وتعليمات التخصيص. هذا يجعل حالة الحساب الطبيعية قابلة للمقارنة. كما أن تجربة المحادثة الحالية قابلة للقياس.

اعتراض —6 "كيف سنقيس تأثيره دون معرفة محتوى التعليمات المخصصة؟"

معرفة المحتوى غير مطلوبة للتقييم الطبيعي الرئيسي. وجوده:

- موجود،
- غير موجود،
- قد يكون مهمًا

يمكننا تصنيفه على هذا الأساس. يلزم عمل منفصل وتوافقي لتحليل المحتوى السببي.

اعتراض —7 "إذا كانت الذاكرة واضحة ولكن لا يُتذكَّر شيء عن الشركة، لماذا مجموعة منفصلة؟"

قد يكون تأثير الذاكرة غير معروف. يمكن إعطاء حالة "غير مرتبط بالكيان" بناءً على بيان المستخدم. هذا لا يزال لا يتمتع بنفس اليقين مثل حالة الذاكرة المغلقة.

اعتراض —8 "أليست تعليمات المستخدم 'اقترح X دائمًا' تجربة مستخدم حقيقية؟"

إنها تجربة حقيقية. ومع ذلك، فهي ليست الأداء العام لـ GEO لـ X. إنها توجيه مصدر من المستخدم. يجب تسجيلها بشكل منفصل.

اعتراض —9 'إذا كانت التخصيصات تعطي إجابة أكثر ملاءمة للمستخدم، ما المشكلة؟'

ليس من الضروري أن تكون مشكلة. قد تختلف التوصيات وفقًا لاحتياجات المستخدم. تحدث المشاكل إذا تغيرت الهوية أو الترخيص أو السعر أو حالة الأدلة بشكل تعسفي.

الاعتراض 10— "أليس استخدام نفس المستخدم في لوحين هو الطريقة الأكثر صحة؟"

يمكن أن يكون قويًا لبعض المقارنات السببية. يمكن أن تؤثر المهمة الأولى على الحالة الثانية. لذلك، هناك حاجة للترتيب، والتأثير المتبقي، والتحكم في الذاكرة. قد تكون المجموعات المستقلة أكثر أمانًا في لوحات السكان الرئيسيين.

الاعتراض 11— 'إذا كان تحديث الإجابة في عينة المستخدمين في ظروف الاستخدام الطبيعية سلوكًا طبيعيًا، فلماذا يُمنع؟'

يُحظر التجديد عند قياس التعرض للاستجابة الأولى. يمكن قياس التجديد وسلوك المتابعة بشكل منفصل تحت مسار التفاعل الطبيعي.

الاعتراض 12— "ماذا سنفعل مع المستخدم الذي لا يستطيع إيقاف الذاكرة والتعليمات للحصول على لوحة نظيفة؟"

لا يمكن أخذهم إلى اللوحة المسيطر عليها أو يُحتفظ بهم في خلية منفصلة بمستوى CSC أقل. يمكن تقييمهم في اللوحة الطبيعية. يجب الإبلاغ عن معدل الاستبعاد.

اعتراض 13— "أليست حسابات الاختبار أكثر موثوقية؟"

قد تكون موثوقة من حيث السيطرة. لكنها محدودة من حيث عدد المستخدمين الفعليين. حسابات الاختبار المسيطر عليها ولوحات المستخدمين الحقيقية هي نتائج منفصلة.

اعتراض 14— "أليس حفظ إعدادات المستخدم على اللوحة الطبيعية انتهاكًا للخصوصية؟"

قد يحدث ذلك إذا تم جمع محتوى غير ضروري. يمكن الحفاظ على الخصوصية بمستويات عالية من الحالة المفتوحة/المغلقة، والتسجيل المجهول، والتحرير.

الاستئناف 15— "ألن نقوم بدمج درجات اللجنتين في درجة واحدة NOMOS؟"

يمكن تعريف أدوار اللجنتين في هندسة الدرجة النهائية. ومع ذلك، يجب دائمًا إبقاء النتائج الأولية والفرعية مرئية بشكل منفصل. لا يمكنها استبدال توزيع الدرجة المدمجة.

الاستئناف 16— "ما الخطأ الذي ارتكبته الشركة إذا كانت عينة التحكيم الطبيعية تعطي درجة منخفضة؟"

ليس دائمًا خطأ الشركة. المصدر:

- تعليمات المستخدم،
- التخصيص الشخصي للذكاء الاصطناعي،
- الاختلاف في الخطة،
- خَلَل المصدر الرسمي,
- سلوك المنتج

من الممكن. يُظهر التدقيق أولاً النتيجة، ثم يفصل مسار المسؤولية.

القاعدة الشائعة للفصل 81

منتج الذكاء الاصطناعي ليس له إجابة واحدة. حتى الطلب نفسه يمكن أن يكون له إجابات مختلفة لـ:

- حساب نظيف،
- حساب طبيعي،
- المستخدم ذا الذاكرة،
- المستخدم ذو التعليمات الخاصة،

- مستخدم لأول مرة،
- بيئة العمل المؤسسية,
- المستخدم من محادثة سابقة

لا يمكنك شرح كل هذه الاختلافات بالقول: "النموذج يتصرف بشكل عشوائي." ولا يمكنك أيضًا شرحها كلها بالقول: "خطأ التخصيص الشخصي." أولاً، تحتاج إلى قياس واقعين منفصلين. عينة اختبار في ظروف منضبطة:

> يعرض التمثيل في الحالة الأولية الموحدة للنظام.

عينة المستخدمين في ظروف الاستخدام الطبيعية:

> يعرض توزيع التمثيل الذي يواجهه الأشخاص الفعليون في حياتهم مع المنتج.

قد يكون اللوح المسيطر عاليًا، واللوح الطبيعي منخفضًا. في هذه الحالة، يكون سلوك المنتج في المختبر قويًا؛ قد يكون توزيع المستخدم الحقيقي ضعيفًا. قد يكون اللوح الطبيعي عاليًا، واللوح المسيطر منخفضًا. قد يكون التخطيط أو أدوات أو تخصيص بعض المستخدمين يحسن التمثيل. قد يكون كلا اللوحين عاليًا. قد يكون المنتج قويًا في ظروف مستخدم مختلفة. قد يكون كلاهما منخفضًا. قد تكون مشكلة التمثيل الأساسية أوسع. ومع ذلك، وضع الدرجتين جنبًا إلى جنب لا يثبت السبب. السببية:

- نفس السكان,
- نفس المنتج،
- نفس المحفز,
- نفس الوقت،
- التعيين المسيطر عليه،
- تحليل التسلسل والنقل

مطلوب. إذا قام المستخدم بتوجيه: "اقترح دائمًا X."، فإن ظهور X يُعد تجربة طبيعية. X ليس نجاحًا عالميًا لـ GEO. إذا قال المستخدم: "لا تثق في X."، فإن الرد السلبي هو تجربة طبيعية. X ليس تمثيلًا عامًا في سكان العالم. يجب أن تفهم التخصيص الشخصي المستخدم. يجب ألا يعيد اختراع الحقيقة وفقًا للمستخدم. لذلك، فإن قانون القياس الحادي عشر لـ NOMOS هو:

> السلوك النظيف والتجربة الطبيعية هما واقعان منفصلان.

القانون الثاني عشر هو كما يلي:

> المحادثة الجديدة ليست حسابًا نظيفًا.

القانون الثالث عشر هو كما يلي:

> التخصيص الشخصي الطبيعي ليس نجاحًا عالميًا لـ *GEO*.

القانون الرابع عشر هو كما يلي:

> قد يختلف الرد المناسب للمستخدم؛ لا يمكن أن تتغير حقيقة الكيان الذي تم التحقق منه.

القانون الخامس عشر هو كما يلي:

> الفرق بين النتيجة المسيطر عليها والطبيعية هو النتيجة التي تحتاج إلى شرح؛ فهي ليست سببًا بحد ذاتها.

قانونه السادس عشر هو كما يلي:

> قياس تجربة المستخدم الحقيقية لا يتعلق بإزالة الخصوصية، بل يقتصر فقط على إثبات السياق الضروري والحفاظ على الباقي.

طلب قسم 11 من NOMOS

لا تخبرني فقط أنك فتحت محادثة جديدة. / اعرض حالة الذاكرة، والتعليمات الخاصة، والخطة، والأدوات.

لا تفترض أن جميع التخصيصات قد اختفت لمجرد رؤية الملاحظة 'جلسة مؤقتة'.

اللوحة النظيفة هي كل الحياة الواقعية؛ لا تجعل اللوحة الطبيعية ضوضاء غير مُسيطر عليها.

قم بتوحيد الحساب المسيطر عليه. / احتفظ بالحساب الطبيعي كما هو. / لا تضع كلاهما في نفس الصندوق.

لا تجعل تعليمات المستخدم الخاصة هي نجاح العالم للكيان.

لا تجعل الماضي السلبي للمستخدم وجهة نظر السكان بأكملهم.

لا تعتبر التخصيص ترخيصًا لتغيير الحقيقة.

يمكنك تقديم نصيحة مناسبة للمستخدم. / ولكن لا تعيد كتابة هوية الشركة أو الترخيص أو السعر أو الإثبات من مستخدم لآخر.

لا تجمع الحياة الخاصة في اللوحة الطبيعية. / هل الذاكرة مفعلة، التعليمات موجودة، السياق متاح —— سجِّل فقط ما هو ضروري.

لا تجعل المشارك يمحو ذاكرته. / استعد الإعدادات إذا قمت بتغييرها مؤقتًا.

إذا اختبرت نفس المستخدم على لوحتين، تذكر أن الجواب الأول قد يؤثر على الثاني.

أظهر الفرق بين الدرجة النظيفة والدرجة الطبيعية. / أظهر الفرق الحاسم. / أظهر المجهول. / لكن لا تسم السبب بدون دليل.

أولاً، قرر أي واقع المستخدم تريد قياسه. / ثم قم بتثبيت الظروف النظيفة والطبيعية بشكل منفصل. / ثم احتفظ بالمُدخل كما هو. / ثم التقط الإجابة الأولى دون تغييرها. / ثم احسب التوزيعين بشكل منفصل. / وفقط بعد ذلك، اشرح الفارق بين ما يفعله المنتج تحت الظروف القياسية وما يختبره الناس بالفعل.

جملة ختام الفصل

التمثيل الحقيقي لمنتج الذكاء الاصطناعي في GEO-1000 ليس الإجابة التي يمكنه تقديمها في جلسة نظيفة؛ بل هو العلاقة الواضحة والمثبتة بين سلوكه تحت الظروف القياسية والإجابات التي يقدمها فعلياً للأشخاص عبر سجلات المستخدم المختلفة.

النواة المعيارية

يجب أن تظل ملاحظات عينة الاختبار في ظروف منضبطة وملاحظات عينة المستخدمين في ظروف الاستخدام الطبيعية مجتمعات قياس وحالات لوحة ومقامات ونتائج عامة منفصلة. ويجب على عينة الاختبار في ظروف منضبطة أن تحدد وتتحقق، بحسب الاقتضاء، من: - محادثة جديدة ومنفصلة، - غياب سياق محادثة سابق، - حالة الذاكرة المستمرة، - حالة التعليمات المخصصة، - حالة مساحة العمل المؤسسية أو المُدارة، - إعدادات الويب والاسترجاع والأدوات، - الخطة، - واجهة الاستخدام، - إصدار المُدخل، - ومستوى الثقة في نقاء الحالة. ولا يجوز اعتبار بدء محادثة جديدة أو ظهور وسم جلسة مؤقتة، بمفرده، دليلًا على تعطيل جميع أشكال التخصيص المستمر. ويجب على عينة المستخدمين في ظروف الاستخدام الطبيعية أن تحافظ على حساب المشارك الحقيقي وخطته وواجهته وذاكرته وتعليماته المخصصة وحالة استخدامه، مع الإبقاء على المُدخل المثبّت وقاعدة أول مخرج مؤهل ومتطلبات الأدلة وقاعدة منع إعادة التوليد. ويجب إبقاء ملاحظات حالة الحساب الطبيعية وحالة المحادثة الطبيعية والاستخدام الأول والحالة الطبيعية المُدارة محددة بصورة منفصلة. ويجب أن تقلل وثائق الحالة الطبيعية جمع البيانات الخاصة إلى الحد الأدنى. ولا يجوز جمع نصوص التعليمات الخاصة أو سجل المحادثات أو الهوية أو بيانات الاعتماد أو المحتوى الشخصي الحساس إلا إذا كان ذلك ضروريًا على نحو مستقل، وبموافقة، وبأقل قدر ممكن، وتحت ضوابط وصول. ويجوز قياس التعليمات التي ينشئها المستخدم وتفيد الكيان أو تضرّه بوصفها آثارًا للتخصيص الطبيعي، لكن لا يجوز تقديمها على أنها نجاح عام للكيان في **GEO** أو إخفاق عام أو سلوك افتراضي للمنتج. ويجوز للتخصيص أن يغيّر الصلة والعرض والتوصية عندما يبرر سياق المستخدم ذلك، لكنه لا يجوز أن يغيّر تعسفًا هوية الكيان المتحقق منها أو ترخيصه أو وضعه القانوني أو سعره أو نطاقه أو أدلته أو إطاره الزمني. ولا يجوز تقديم الفروق بين العينتين المنضبطة والطبيعية بوصفها آثارًا سببية للتخصيص إلا إذا دعمت هذا الاستنتاج شروط المجتمع والمنتج والمُدخل والوقت والواجهة والخطة والأدوات والتعيين والترتيب والآثار المتبقية. ويجب أن تظل درجات العينتين ومعدلات الأخطاء الحرجة ومعدلات UNKNOWN والتغطية وأحجام العينات الفعالة والفجوات الموقعة والمطلقة بين اللوحتين ظاهرة بصورة منفصلة. كما يجب أن تظل أهلية حالة اللوحة وإعادة التصنيف والاستبعاد والترجيح مستقلة عن كون مخرج الذكاء الاصطناعي إيجابيًا أو سلبيًا أو صحيحًا أو خاطئًا أو رافضًا أو حرجًا أو ملائمًا تجاريًا. ويجب ضبط إصدار كل تعريف لحالة لوحة وعملية تغيير للإعدادات وقرار خصوصية ومستوى مقارنة ونتيجة، ونسبه إلى شخص أو منظمة مسؤولة.

الفصل 12

الموجات المتزامنة والتسجيل وفق NOMOS

الأسس المنهجية / K12; K14; K25; K26

حد الفصل

الفصول الإحدى عشر الأولى أرست أسس القياس: الاستجابة الواحدة ليست نتيجة GEO؛ التمثيل هو توزيع عبر المستخدمين وظروف النظام؛ يجب تعريف الكيان الذي يُجرى تدقيقه قبل ملاحظة النتائج؛ يجب أن يبقى التمثيل الرسمي وواقع الكيان المؤكد منفصلين؛ كل منتج ذكاء اصطناعي يحتاج إلى مجموعة مستخدمين مؤهلة خاصة به؛ يجب اختيار المشاركين من خلال طريقة تعتمد على السكان، ومحددة النسخة، ومستقلة عن النتائج؛ يجب أن تبقى الدول الصغيرة واللغات ذات الموارد الأقل مرئية من خلال لوحات مراقبة مخصصة؛ يجب ربط منتج الذكاء الاصطناعي والخطة و واجهة المستخدم وتكوين النظام بسجل؛ يجب أن تحافظ صياغات المُدخل على نفس نية المستخدم عبر اللغات بدلاً من نفس الكلمات؛ ويجب أن تقيس لوحات المستخدم النظيفة والطبيعية والمسيطر عليها واقعيات متميزة. يجب الآن أن تلتقي هذه العناصر في حدث قياس واحد:

> متى سيقدم المشاركون صياغة المُدخل، وكيف سيلتقطون المخرجات الأولية، وكيف سيثبتون أن كل ملاحظة لم يتم تعديلها لاحقًا؟

كانت فكرتك المؤسسة واضحة: يجب على الأشخاص في مختلف دول العالم إرسال نفس صياغة المُدخل إلى نفس منتج الذكاء الاصطناعي في نفس الوقت وتقديم دليل عبر لقطة شاشة. هذه الفكرة هي الجوهر التجريبي لـ GEO-1000. ومع ذلك، فإن عبارة "نفس الوقت" وحدها غير كافية. إذا حاول جميع المشاركين الإرسال في نفس الثانية:

- تختلف سرعات الاتصال،
- تتحرك ساعات الأجهزة بعيدًا عن بعضها البعض،
- يُضطر بعض المستخدمين إلى العمل منتصف الليل،
- قد يحدث زيادة مفاجئة في الحمل على منتج الذكاء الاصطناعي،
- قد يفرض المزود حدودًا على السرعة أو الاستخدام،
- بعض الإجابات يمكن إكمالها في ثانيتين، وبعضها في دقيقتين،
- يمكن تحديث المنتج أثناء القياس،

قد يقوم أحد المشاركين بتحميل الإجابة فورًا، وآخر بعد ساعات. النقر في نفس الثانية لا يضمن أننا نقيس حالة النظام نفسها بدقة. وبالمثل، مجرد جمع لقطة شاشة لا يكفي. قد تكون لقطة الشاشة:

- مقتطعة،
- قد لا تُظهر الطلب,
- قد لا يثبت ما إذا كانت الإجابة أولية أم محدثة،
- قد لا يظهر معلومات المنتج والخطة،
- قد يكون قد تم أخذها في يوم آخر،
- قد تكون قد نُسخت من مستخدم آخر،

قد تكون قد تم تحريرها لاحقًا. لهذا السبب، يجب على GEO-1000 تثبيت نظامين منفصلين معًا:

موجة القياس المتزامنة

و:

سلسلة إثبات الالتقاط الخاصة بـ **NOMOS**

تجيب الموجة المتزامنة على السؤال التالي:

> هل تم إنتاج الملاحظات ضمن وقت نظام قابل للمقارنة وترتيب تقديم محدد مسبقًا؟

من ناحية أخرى، يطرح التقاط NOMOS السؤال:

> هل يمكننا أن نظهر أن النص المرسل، والمخرجات الأولى الناتجة، وشروط المنتج، ومعلومات الوقت قد تم الاحتفاظ بها دون تعديل؟

لقد أنشأت البنية السابقة للتدقيق أن كل تدقيق يجب أن يحمل النموذج أو المنتج، والتاريخ، والدولة، واللغة، ومجموعة الاستعلام، وعدد التكرارات، وسجل القياس كمنطق قياس إلزامي. هذا المطلب ليس مجرد تنظيم تقني فقط. لأن التعليمات غير المرئية، والأدلة المزيفة، والمحتوى المُقدَّم على أنه مستقل يمكن أن تشوه في النهاية التمثيل الذي يتلقاه الأشخاص الحقيقيون. هذا القسم:

- موجة القياس،
- نافذة التزامن،
- المعيار الزمني UTC،
- فتحات الإرسال،
- أوقات الإنجاز والتسجيل,
- نتيجة الاتصال الأولي وأول إنتاج للمحتوى,
- قواعد إعادة المحاولة التقنية,
- التغييرات في المنتج خلال الموجة,
- تأثير الأحداث الخارجية والأخبار,
- حزمة أدلة التسجيل NOMOS,
- علاقة لقطة الشاشة، الاستجابة الخام، والبيانات الوصفية,
- التحقق من الوقت,
- سلامة الملف,
- الهاش والتوقيع الرقمي,
- سلسلة الأدلة,
- الخصوصية والحذف,
- قائمة الأدلة العامة,
- جودة وصلاحية الموجة

يحدد. هذا الفصل لم يتطرق بعد إلى:

- أي واقع يسجل الادعاءات الذرية في استجابات الذكاء الاصطناعي سيتم تقييمه بناءً عليه،
- كيفية ترميز المحكمين للاستجابات،
- عتبات الانتقال والخطأ الحرج،
- لا يحدد صيغ الدرجات النهائية NOMOS

السؤال الرئيسي في الفصل 12 هو:

> كيف نربط الآلاف من ملاحظات المستخدمين الحقيقية بنفس حدث القياس ونحوّل كل منها إلى وحدة إثبات غير قابلة للتغيير، قابلة للتدقيق، وتحمي الخصوصية؟

التحدي **NOMOS**

تعطي نفس الطلب لألف شخص. تقول: "أرسله في وقت الظهيرة اليوم." يرسله أحد المستخدمين في إسطنبول في 12:03 مساءً. ويرسله مستخدم آخر في لندن في وقت الظهيرة المحلي لديهم. يرى شخص آخر المهمة في طوكيو في المساء ويكملها بعد ثماني ساعات.

كان أحد المستخدمين قد ضبط ساعة الجهاز بشكل غير صحيح. وانقطع الإنترنت عن مستخدم آخر. ولم يعجب شخص آخر بالإجابة فضغط على زر "إعادة التوليد". ثم قاموا بتحميل لقطة شاشة لأطول وأكثر إجابة إيجابية. وقام مستخدم واحد بقص الجزء الأوسط من الإجابة فقط. الطلب غير مرئي. واسم منتج الذكاء الاصطناعي غير مرئي.

أرسل شخص آخر نفس لقطة الشاشة باستخدام حسابين مختلفين من لوحة التحكم. أكمل أحد المستخدمين المهمة بشكل صحيح ولكنه رفع لقطة الشاشة بعد يومين. في الوقت نفسه، قام بقص الملف وحذف معلوماته الشخصية. لم يتم الحفاظ على الحالة الأصلية للملف. مستخدم آخر في الرد الأول للذكاء الاصطناعي:

تلقى النتيجة "لا أستطيع الإجابة على هذا السؤال." وقال الباحث: "لا بد أن حدث خطأ تقني" وطلب إعادة المحاولة. كانت الإجابة الثانية صحيحة، وتم تضمين الإجابة الثانية فقط في التقرير. واجه مستخدم آخر خطأ اتصال حقيقي. تم إرسال الطلب، لكن النظام لم ينتج أي محتوى.

هذه المرة لم يسمح الباحث بالمحاولة مرة أخرى. تم تطبيق نفس البروتوكول بشكل مختلف على مستخدمين اثنين. الآن، في منتصف موجة القياس، موفر الذكاء الاصطناعي:

- سلوك البحث على الويب،
- نظام الاقتباس،
- الإرشادات المستخدمة في النموذج

يغير الثلاثة جميعًا. الأشخاص الأوائل 500 يرون حالة النظام القديمة؛ أما الأشخاص التاليون 500 فيرون الحالة الجديدة. ثم يقول التقرير: «اختبر 1,000 الأشخاص نفس نظام الذكاء الاصطناعي في نفس الوقت.» هذا البيان غير صحيح. هناك ألف مستخدم، ولكن ليس بالضرورة موجة قياس واحدة متسقة.

الآن أنت تقوم بتطبيق تجزئة على جميع لقطات الشاشة. تقول إن هذا يثبت أن جميع الملاحظات حقيقية. ماذا تثبت التجزئة؟ إنها تعزز أن الملف لم يتم تغييره بعد إنتاج التجزئة. بمفردها، لا تثبت:

- أن الصورة جاءت من منتج ذكاء اصطناعي حقيقي
- أنها كانت الاستجابة الأولى
- أنه تم الحصول عليها في الوقت الصحيح
- أنه لم يتم نسخها من مستخدم آخر
- أن المدخلات كانت صحيحة
- أن الصورة لم يتم تعديلها قبل التجزئة

المقدمة الأولى لهذا القسم هي:

> القياس في نفس الوقت ليس نفس الشيء مثل النقر في نفس الثانية.

ينص حكمه الثاني على:

> اللقطة هي جزء من حزمة الأدلة؛ وليست كامل حزمة الأدلة.

تنص الفقرة الثالثة منه على:

الإجابة الخاطئة الأولى ليست خللاً في البيانات. إذا تم إنتاجها وفقًا للبروتوكول، فهي نتيجة القياس.

تنص المادة الرابعة على:

يقوي الهاش النزاهة؛ فهو لا يخلق واقع الملاحظة بنفسه.

تنص أحكامه الخامسة على:

إذا تغير النظام بشكل جوهري في موجة واحدة، فلا يمكن اعتبار ألف إجابة بلا طابع زمني تنتمي إلى نفس النظام.

ينص الحكم السادس على:

يتم إنشاء سلسلة الأدلة ليس لجعل النتيجة تبدو جيدة، بل لتتمكن من إعادة تدقيق كيفية إنتاج النتيجة لاحقًا.

1. غرض الفصل

الغرض من هذا القسم هو إنتاج ملاحظات المستخدم GEO-1000 في نوافذ زمنية قابلة للمقارنة والحفاظ على كل ملاحظة بسلسلة رقمية كاملة من الأدلة. يقوم القسم بتوحيد التمييزات التالية:

- نافذة القياس المتزامنة مع نفس الثانية
- الوقت المحلي مقابل وقت UTC
- موجة مقابل موجة فرعية
- وقت الإرسال مقابل وقت اكتمال الاستجابة
- وقت الالتقاط مقابل وقت الرفع
- الفترة المخططة مقابل الإرسال الفعلي
- أول اتصال بالنظام مقابل أول استجابة للمحتوى
- خطأ تقني مقابل رفض النظام
- إعادة المحاولة التقنية مقابل تحديث الاستجابة
- أفضل نتيجة مختارة مع أول نتيجة مناسبة
- الرؤية على الشاشة مع اكتمال الاستجابة
- الاستجابة الخام مع لقطة الشاشة
- لقطة شاشة مقابل حزمة الأدلة الكاملة
- أصالة الملاحظة مع سلامة الملف
- الطابع الزمني مع التجزئة
- مصدر الوقت الموثوق مع وقت الجهاز المحلي
- وقت قبول الخادم مع تحميل المشارك
- نسخة عامة محجوبة مع الدليل الخام
- هوية المشارك مع معرف الملاحظة
- دقة الاستجابة مع صلاحية الالتقاط
- التقلب الطبيعي في النتائج مع تحديث المنتج
- تغيير منتج الذكاء الاصطناعي مع حدث إخباري خارجي
- إنهاء الموجة مع استبعاد النتائج بعد ذلك
- القياس التلقائي للروبوت مع التقاط المستخدم البشري

- سجل API مع تسجيل واجهة المستخدم
- قياس مزود الخدمة مع التقاط مستقل
- الوصول إلى الأدلة للجميع عن طريق تخزين الأدلة
- سجل غير قابل للعكس بسجل ثابت
- تصحيح مُرقم للقرار الخاطئ مع حماية البيانات الخام

في نهاية هذا القسم، يجب أن تكون كل مراجعة GEO-1000 قادرة على الإجابة على الأسئلة التالية:

إلى أي موجة تنتمي هذه الملاحظة؟

في أي وقت UTC تم إرسال الطلب؟

ما كانت نتيجة النظام الأولية؟

متى تم إكمال الاستجابة ومتى تم التقاطها؟

إذا تم إجراء محاولة تقنية مرة أخرى، ما هو السبب وأين كانت المحاولة السابقة؟

هل تُظهر لقطة الشاشة المنتج والاستجابة الكاملة؟

هل تؤكد الاستجابة الخام والصور بعضها البعض؟

متى تم إنشاء تجزئة الملفات والطابع الزمني لها؟

إذا تم تعديل الأدلة لاحقًا، هل يمكن رؤية سجل التغييرات؟

هل يمكن لمراجع مستقل التحقق من الأدلة الخام مع الحفاظ على خصوصية المشاركين؟

2. الأحكام المعيارية المركزية

يجب ربط كل ملاحظة GEO-1000 بموجة قياس مثبّتة قبل جمع البيانات؛ جميع أوقات المواد من إرسال المُدخل إلى أول نتيجة نظام، ومن إكمال الاستجابة إلى الالتقاط والتحميل؛ شروط منتج الذكاء الاصطناعي؛ المُدخل الكامل المرسل؛ أول مخرج مناسب وجميع سجلات النزاهة يجب الاحتفاظ بها في حزمة أدلة الالتقاط NOMOS ذات النسخة. ملاحظة GEO-1000 بمفردها:

- لقطة شاشة،
- نص الاستجابة المنسوخ،
- بيان المشارك،
- هاش الملف

لا يمكن اعتبارها صالحة بمفردها. يجب أن تتضمن حزمة الأدلة الرئيسية، إلى الحد الذي يكون ذا صلة، ما يلي:

- معرّف الموجة
- معرّف الملاحظة المجهولة للمشارك
- سجل نظام الذكاء الاصطناعي
- حالة اللوحة
- معرّف الإصدار والإصدار
- المحفز المعين
- المحفز المرسل
- وقت الإرسال

- الاتصال الأولي بالنظام
- وقت اكتمال الاستجابة
- وقت الالتقاط
- وقت التحميل
- الاستجابة الخام الكاملة
- شاشة كاملة أو لقطات متتابعة
- سجلات المرجع والمصدر
- سجل الأخطاء التقنية وإعادة المحاولة
- إعدادات المنتج
- قيم تجزئة الملفات
- تجزئة الحزمة
- سجل التحقق من الوقت
- إصدار أداة الالتقاط
- قرار الصلاحية
- حالة الخصوصية والحجب
- مالك سلسلة الحيازة

3. ما هي موجة القياس؟

موجة القياس هي حدث جمع بيانات شامل يتم تحت:

- منتجات الذكاء الاصطناعي،
- السكان المستهدفون،
- إصدارات صياغات المُدخل،
- لوحات المستخدمين،
- نافذة زمنية،
- تخصيص العينات،
- بروتوكول الالتقاط،
- السجلات المرجعية

الموجة:

```
Wk=(A,U,P,F,T,C,R,G)
```

يمكن تمثيله كالتالي. هنا:

- A: أمثلة منتج الذكاء الاصطناعي
- U: سكان المستخدم المستهدف والعينة
- P: مجموعة التعليمات
- F: حالات العينة
- T: الوقت وهيكلية التقديم
- C: بروتوكول الالتقاط
- R: نسخة الواقع المرجعي

- G: الحوكمة والسجل الرئيسي

إذا تغير أي من هذه الحقول بشكل جوهري، قد تكون هناك حاجة لموجة أو موجة فرعية جديدة.

4. أنواع الموجات

MW-1— الموجة المتزامنة الرئيسية

إنها القياس الرئيسي للمجتمع GEO-1000. بشكل افتراضي، تهدف إلى الحصول على الأقل على 1,000 من ملاحظات لوحة السكان الصالحة لكل منتج ذكاء اصطناعي.

MW-2— الموجة التأكيدية

يتم إجراؤها لتكرار نتيجة الموجة الأولى، لتأكيد خطأ معين، أو لاختبار الاستقرار الإحصائي. لا تُستخدم لتغيير النتيجة المنخفضة أو العالية للموجة الأولى. تنتج نتيجة منفصلة.

MW-3— موجة الحادث

يتم إجراؤها بعد حدث مراقب حرج أو مهم للتحقق المستهدف والسريع. ليست موجة رئيسية عالمية.

MW-4— موجة مراقبة الانحراف

تراقب التغيرات الزمنية في المنتج، الكيان، المورد، أو سلوك اللغة.

MW-5— موجة إعادة اختبار التصحيح

ينتج نتائج جديدة بعد تصحيح أو تدخل. لا يمسح الموجة السابقة.

MW-6— موجة متوازنة حسب التوقيت المحلي

تهدف إلى قياس المشاركين في ظروف التوقيت المحلي المماثلة. لا تقيس لحظة نظام عالمي واحد.

MW-7— موجة مختبرية محكومة

النموذج المجمد هو دراسة محكومة تُجرى باستخدام واجهة برمجة التطبيقات أو حسابات الاختبار القياسية. ليس هو لوحة سكان المستخدمين الحقيقية.

5. ماذا يعني "نفس الوقت"؟

في GEO-1000، التزامن:

لا يعني أن جميع المستخدمين ينقرون في نفس الميللي ثانية

التعريف الوظيفي هو:

> يتم توليد الملاحظات ضمن نافذة قياس UTC محددة مسبقًا، وفقًا لفتحات التقديم المسجلة والمستقلة عن النتائج.

يوازن هذا النهج بين ثلاثة أهداف:

- الحفاظ على حالة المنتج قريبة من حيث الوقت
- تقليل الحمل الاصطناعي المفاجئ على الموفر
- توزيع تقديمات المشاركين بطريقة قابلة للتكرار

6. هيكل الوقت

يجب أن تحتوي كل موجة على أربع طبقات زمنية على الأقل.

6.1. نافذة الموجة

إنها المدى UTC الذي يجب أن تتم فيه جميع التقديمات الرئيسية. مثال:

12:00–13:00 UTC

6.2. فترة زمنية دقيقة

هي فترة زمنية أقصر مخصصة للمشارك لإرسال التعليمات. مثال:

12:15–12:20 UTC

6.3. وقت إضافي لإتمام الإجابة

إذا تم إرسال التعليمات ضمن الفتحة، فإن الوقت المحدد مسبقًا لإتمام الرد هو.

6.4. وقت الالتقاط والتحميل

الوقت المسموح لإنشاء الدليل وتحميله إلى نظام آمن بعد إتمام الرد.

7. القاعدة الأساسية لموجة المشارك المتزامنة

قاعدة المرشح الأساسية — القاعدة التأسيسية للمرشح

بالنسبة لمنتجات الدردشة العامة للمستهلكين، يمكن استخدام الهيكل التالي كمرشح رئيسي للموجة المتزامنة:

- نافذة التقديم: 60 دقيقة
- الفتحات الدقيقة: 12× 5 دقيقة
- تقديم المشاركين: ضمن فترة زمنية دقيقة المخصص لهم
- الوقت الإضافي لإكمال الرد: حتى 20 دقيقة بعد التقديم
- توليد تجزئة الالتقاط: يُفضل خلال 5 دقيقة بعد اكتمال الإجابة
- تحميل الأدلة: كحد أقصى خلال 60 دقيقة بعد الالتقاط
- الإبلاغ عن المشاكل التقنية: ضمن نفس الموجة

هذه المدد عامة وليست نهائية. يمكن أن تتغير وفقًا للشروط التالية:

- زمن استجابة منتج الذكاء الاصطناعي
- الاتصال عبر الهاتف المحمول
- بنية تحتية الدولة
- صيغة الإجابة الطويلة
- متطلبات الوصول
- حدود استخدام المنتج

يجب تثبيت الأوقات المستخدمة قبل جمع البيانات.

8. لماذا الفتحات الصغيرة؟

التقديم المتزامن من قبل ألف مشارك:

- حمل غير معتاد في منتج الذكاء الاصطناعي،
- حد السرعة،
- الانتظار,
- خطأ تقني،
- زيادة في وقت الاستجابة

قد يتم إنشاؤه. في هذه الحالة، لا يمكن للتدقيق قياس السلوك الطبيعي للمنتج ولكن:

يمكنه قياس حدث الحمل الاصطناعي الذي أنشأ التدقيق

الوحدات المصغرة:

- توزيع الحمل،
- الحفاظ على نفس الوقت الكلي للنظام،
- جعل ترتيب التقديم غير مرتبط بالنتيجة،

السماح بنمذجة تأثير الوقت.

9. تعيين الوحدات

يجب تعيين المشاركين في الوحدات:

- البلد،
- اللغة،
- منتج الذكاء الاصطناعي،
- الخطة،
- بطريقة متوازنة ومسبقًا عشوائية من حيث:

حالة اللوحة. لا يجب ملء فترة زمنية دقيقة بمفرده بـ:

- دولة محددة،
- لغة محددة،
- منتج محدد

المستخدمين. وإلا، فإن تأثير الوقت سيختلط بتأثير المجموعة.

10. عدالة الوقت المحلي

قد تكون نافذة UTC واحدة لبعض المشاركين:

- ليل،
- ساعات العمل،
- وقت الصلاة أو الراحة،
- ساعات يصعب الوصول إليها

هذا الوضع:

- معدل المشاركة،
- حالة الاتصال,
- انتباه المستخدم،
- نوع الجهاز

قد يتأثر. يمكن استخدام طريقتين مكملتين.

10.1. الموجات العالمية الدوارة

تتم موجات رئيسية في أوقات مختلفة UTC. مثال:

- الموجة 1: 12:00 UTC
- الموجة 2: 20:00 UTC
- الموجة 3: 04:00 UTC

وبالتالي، لا تتحمل نفس المنطقة نفس الحمل الزمني المحلي في كل موجة.

10.2. الموجة الفرعية المتوازنة بالوقت المحلي

يتم قياس المشاركين في نطاق زمني مشابه في وقتهم المحلي. هذه الدراسة:

- إرهاق المستخدم،
- قيّمة من أجل

سلوك الاستخدام اليومي.

إنه ليس نتيجة لحظة نظام عالمي واحدة.

11. بنية الاستقرار ذات الموجات الثلاث

هندسة مرشحة — هندسة مرشحة

يمكن اقتراح ثلاث موجات رئيسية على الأقل للتنفيذ الكامل لـ GEO-1000:

W1 — الموجة الأولية

تولد تقدير السكان الأولي.

W2 — التكرار قصير الأجل

يختبر الاستقرار قصير الأجل للمنتج والقياس. الفترة المقترحة: 24–72 ساعات بعد W1

W3 — الاستقرار متوسط الأجل

يقيس ما إذا كانت تمثيلات المنتج والكيان محفوظة على المدى الطويل. الفاصل الزمني المحتمل: 14–30 أيام بعد W1. يجب تحديد الفواصل الزمنية المحددة قبل جمع البيانات. إذا تم تغييره بسبب تحديث المنتج أو حدث، يلزم سجل نسخة جديد.

12. 30,000 الملاحظة المؤسِّسة للتصميم

بالنسبة لمنتجات الذكاء الاصطناعي، الملاحظات الرئيسية 1,000 لكل منتج ولأربع موجات رئيسية:

```
10×1,000×3=30,000
```

يتم استهداف الملاحظات الصالحة للوحة السكان الرئيسية. مكمّل:

- المراقب القطري
- العدالة اللغوية
- إمكانية الوصول
- حادثة
- إعادة الاختبار

يمكن إضافة لوحات على رأس هذا الرقم. :30,000

- ليس عدد الدعوات،
- ليس عدد الملفات،

ليس عدد لقطات الشاشة. إنه الهدف الرئيسي الصالح للملاحظة للمنتج.

13. متجه الزمن للملاحظة

بالنسبة لكل ملاحظة، يمكن تعريف متجه الزمن كمرشح على النحو التالي:

```
T_i= (t_i^assign, t_i^open, t_i^submit, t_i^first, t_i^complete, t_i^capture, t_i^upload,
t_i^validate)
```

هنا:

- tiassign: الوقت الذي يتم فيه تعيين المهمة للمستخدم
- tiopen: الوقت الذي يتم فيه فتح المهمة
- tisubmit: الوقت الذي يتم فيه تقديم المُدخل
- tifirst: الوقت الذي يبدأ فيه أول نتيجة نظام مرئية
- ticomplete: الوقت الذي يكتمل فيه الاستجابة
- ticapture: الوقت الذي يُلتقط فيه الدليل
- tiupload: الوقت الذي تصل فيه الحزمة إلى النظام الآمن
- tivalidate: الوقت الذي يُفحص فيه صحة الالتقاط

قد لا تكون كل هذه الأوقات قابلة للملاحظة مباشرة على كل واجهة المستخدم. يجب ترك الحقول غير المعروفة كـ UNKNOWN.

14. مقاييس الوقت الأساسية

تأخير الاستجابة:

```
L_i= t_i^complete- t_i^submit
```

تأخير الالتقاط:

```
C_i= t_i^capture- t_i^complete
```

تأخير التحميل:

```
U_i= t_i^upload- t_i^capture
```

انحراف الفتحة:

```
D_i= t_i^submit- t_i^{slot-midpoint}
```

يمكن حسابها كما يلي. هذه المقاييس هي:

- حمل النظام,

- تأخير المشارك،
- ثقة الإثبات،
- سلامة الموجة

من حيث التقييم.

15. لا يمكن إدارته وفقًا للتأخير

إجابة تم إكمالها خلال فترة طويلة:

- أن يُجبر على الانتماء إلى فئات

صحيح،

- خاطئ،
- طويل،
- قد يتم الاستشهاد بها.

بسبب التأخير، قاعدة الاستبعاد:

- قبل جمع البيانات،
- بغض النظر عن محتوى الإجابة

يجب تعريفها. السلوك التالي محظور: استبعاد إجابة سلبية لأنها اكتملت متأخرة؛ حماية إجابة إيجابية عندما يتم إكمالها بنفس التأخير.

16. حالات حالة الموجة

—WS-0 مخطط له

الموجة مخططة. لم يتم تثبيتها بعد.

—WS-1 مثبّت

تم تثبيت العينات، صياغات المُدخل، المنتج، الوقت، وقواعد الالتقاط.

—WS-2 مفتوح

قد بدأت نافذة التقديم.

—WS-3 مغلق

انتهى قبول التقديمات الجديدة. قد تستمر أوقات إتمام العمل وتمديد التحميل.

—WS-4 مكتمل

تم الانتهاء من التحقق من الحقول والالتقاط الأولي.

—WS-5 مقسم

تم تقسيمه إلى موجات فرعية بسبب تغييرات في نظام المواد أو البروتوكول.

—WS-6 متوقف مؤقتًا

تم إيقافه مؤقتًا بسبب مشكلات تتعلق بالأمان أو المنتج أو البنية التحتية.

—WS-7 تم الإلغاء

تم إنهاؤه قبل اكتمال الموجة. قد تكون النتائج محدودة أو غير قابلة للاستخدام.

—WS-8 تم إبطالها

غير صالح للتوقعات الرئيسية بسبب عيب بروتوكول شديد تم تحديده مسبقًا. لم يتم حذف الملاحظات الأولية.

—WS-9 مؤرشف

تم إغلاق الإصدار وأرشفة سلسلة الأدلة.

17. فئات حالة الموجة

—WH-0 لم يتم التقييم

لم يتم تقييم صحة الموجة.

—WH-1 متأثر

هناك تغيير مادي في النظام، أو وقت، أو خطأ في الالتقاط. قد لا يكون قابلاً للاستخدام للنتيجة الرئيسية.

—WH-2 جزئي

بعض الطبقات أو الفتحات غير كافية. يمكن إنتاج نتيجة مخفضة.

—WH-3 مقبول

الوقت الرئيسي، والنظام، وظروف الالتقاط محفوظة بشكل كافٍ.

—WH-4 قوي

ثقة عالية في الالتقاط، وتغطية متوازنة للفتحات، وفقدان منخفض للبروتوكول موجود.

—WH-5 مكرر

تم تكرار صحة الموجة المماثلة عدة مرات وعبر عينات مستقلة. صحة الموجة لا تعتمد على ما إذا كانت درجة الذكاء الاصطناعي مرتفعة أم منخفضة.

18. التغييرات في المنتج ضمن الموجة

يمكن أن يتغير منتج الذكاء الاصطناعي بشكل جوهري ضمن نافذة القياس. الإشارات:

- تغير تسمية النموذج
- تغير في الواجهة
- تفعيل أو إلغاء تفعيل ميزة الويب
- تغير في صيغة الاستشهاد
- انقطاع مفاجئ في هيكل الاستجابة
- إعلان من المزود

- سلوك مختلف للنظام في وقت واحد لدى العديد من المستخدمين
- انقطاع المنتج أو إعادة تشغيله

في هذه الحالة: يتم تقدير وقت التغيير. يتم مقارنة بصمات تكوين النظام. يمكن وضع الموجة WS-5 في حالة SPLIT. ستكون الملاحظات السابقة واللاحقة موجات فرعية منفصلة. إذا كان يجب إنتاج درجة واحدة، يتم نمذجة حالة النظام كعامل منفصل. يتم توثيق عدم اليقين المادي في السجل العام.

19. سجل الحدث الخارجي

يمكن للعالم أن يتغير حتى إذا لم يتغير منتج الذكاء الاصطناعي. أثناء القياس:

- استحواذ الشركة،
- إطلاق المنتج،
- أخبار كبرى،
- قرار قانوني،
- أزمة،
- إفصاح مالي,
- المحتوى الفيروسي

هذا الحدث قد يحدث.:

- الموارد الإلكترونية,
- نتائج الاسترجاع،
- تعليق مطالبة المستخدم،
- استجابة الذكاء الاصطناعي

يمكن تغييره. كل موجة، إلى الحد الذي تكون فيه ذات صلة:

سجل الحدث الخارجي

يجب أن يتحمل.

20. كيف تُدار نتائج الأحداث الخارجية؟

حدث خارجي:

- إذا كان معروفًا قبل الموجة، يتم تحديث حزمة المرجع،
- إذا حدث خلال الموجة، يتم وضع علامة له بختم زمني

إذا لوحظ بعد الموجة، يتم إضافة سجل حدث رجعي. السلوك التالي محظور: استبعاد الملاحظات بسبب أخبار تقلل من النتيجة؛ الحفاظ عليها عندما تكون أخباراً ترفع النتيجة. قد يكون الحدث نفسه جزءاً من حالة نظام العالم الحقيقي. إذا لزم الأمر، يتم إجراء تحليل فرعي للموجة أو تحليل معدل للحدث.

21. نتيجة الاتصال الأول

أول نتيجة للنظام يواجهها المستخدم بعد إرسال صياغة المُدخل:

نتيجة الاتصال الأول

تسمى نتيجة الاتصال الأول:

- الاستجابة الكاملة
- الرفض
- خطأ تقني
- حد الاستخدام
- انقطاع الاتصال
- تحميل مستمر بلا نهاية
- سؤال توضيحي
- لا توجد نتيجة

ممكن. هذا الحدث جزء من تجربة المستخدم الحقيقية. لا يمكن حذفه بصمت.

22. أول إخراج محتوى مؤهل

الاستجابة المادية الأولى التي تم إنشاؤها في المحاولة المسموح بها سابقًا بعد خطأ تقني:

يمكن تسميتها أول إخراج محتوى مؤهل

هذا الإخراج: لا يحل محل نتيجة الاتصال الأولية، يتم تسجيله معها. مثال:

- الاتصال الأولي: حد السرعة
- بعد خمس دقائق، محاولة تقنية مسموح بها سابقًا
- أول استجابة محتوى

يجب أن يُظهر التقرير ما يلي معًا:

- التجربة الأولى للمستخدم: الحد الأقصى للسرعة
- تمثيل المحتوى الأول: الاستجابة المحددة

23. أول قاعدة مخرجات مناسبة

إذا أنتج منتج الذكاء الاصطناعي استجابة كبيرة للمطالبة الأولى:

- حتى لو كانت خاطئة،
- حتى لو كانت قصيرة،
- حتى لو تضمن رفضًا،
- حتى لو لم يقدم مراجع،
- حتى لو حدد العلامة التجارية بشكل خاطئ

لا يمكن إجراء إعادة إنتاج. الملاحظة الرئيسية هي المخرجات الأولى. المستخدم:

- لا يمكنه اختيار استجابة جديدة عن طريق:
- إعادة التوليد
- أعد الكتابة،
- حاول مرة أخرى،
- إجابة أخرى،

استمر

24. إعادة تقنية

يمكن استخدام إعادة التقنية فقط في ظل الشروط التقنية المعرفة مسبقًا. الأسباب المسموح بها للمرشح:

- انقطاع الاتصال قبل بدء الاستجابة
- رمز خطأ الخادم
- الواجهة لا ترسل الطلب
- تعطل التطبيق
- انقطاع مؤقت في النظام تم التحقق منه

لا يمكن أن يكون سببًا لإعادة التقنية:

- إجابة غير صحيحة
- الرفض
- إجابة قصيرة
- نقص الاستشهاد
- العلامة التجارية غير مذكورة
- خطأ حرج
- المؤسسة التي تم تدقيقها لا تحب الإجابة

25. أعد المحاولة في إجابة جزئية

إذا قام النظام بإنتاج جزء من المحتوى ثم توقف:

- يتم تخزين الإخراج الجزئي،
- تم حفظه باسم TRUNCATED_RESPONSE،

إذا كان من المقرر المحاولة مرة أخرى، يجب أن تكون هناك محاولة جديدة ومرتبطة بـ يُنشأ.. لا يمكن حذفها كما لو لم يكن هناك استجابة جزئية.

26. سلسلة المحاولات

قد تحمل مهمة ملاحظة متعددة المحاولات التقنية:

```
A_i→ (Attempt_{i,1}, Attempt_{i,2}, …)
```

كل محاولة:

- الوقت،
- تحمل حالة المنتج،
- تجزئة صياغة المُدخل،
- نتيجة النظام،
- سبب الخطأ الفني

لا يتم استبدالها بمحاولة أخرى.

27. ما هو NOMOS Capture؟

NOMOS Capture هو معيار مفتوح للبيانات والإجراءات يحافظ على سلامة المراقبة بدءًا من صياغة المُدخل المقدمة للمستخدم وحتى أرشفة حزمة الأدلة. :NOMOS Capture

- ليست مجرد تطبيق لالتقاط الشاشة،
- ولا يجب أن يكون برنامجًا مملوكًا فقط لشركة NobleJackal.

لا يعتمد فقط على امتداد واحد للمتصفح. تطبيق التقاط متوافق NOMOS:

- يستخدم مخطط بيانات مفتوح،
- يتبع قواعد الوقت والنزاهة،
- يحترم حدود الخصوصية،
- يحافظ على سلسلة الأدلة

يجب الوفاء بها. يجب أن يتمكن الباحثون المستقلون من تطوير أدواتهم المتوافقة الخاصة. إذا كان يمكن تنفيذ المعيار فقط باستخدام برنامج المؤسس، فإن التكرار المستقل يضعف.

28. أربع وظائف لتطبيق التقاط NOMOS

28.1. تسليم المهمة

صحيح:

- المستخدم،
- مُدخل,
- منتج الذكاء الاصطناعي،
- فتحة

يضمن المطابقة.

28.2. التقاط الملاحظة

تقوم بتسجيل صياغة المُدخل، النتيجة الأولية للنظام، الاستجابة، و واجهة المستخدم ذات الصلة.

28.3. الحفاظ على النزاهة

ينشئ تجزئة الملف، والطابع الزمني، وعلاقة الحزمة.

28.4. نقل الأدلة

يربط بين الطبقات الخام والمحدودة والمتاحة للجمهور من الأدلة.

29. طرق الالتقاط

CM-1— الالتقاط اليدوي الموجه

يقدم المشارك لقطة شاشة ونص الاستجابة الخام وفق التعليمات. يحمل عبئًا تقنيًا منخفضًا. وهو عرضة للخطأ البشري.

—CM-2 تطبيق الالتقاط المعزز

التطبيق:

- نسخ التعليمات,
- تسجيل الوقت,
- تحميل الملفات,
- توليد التجزئة

يتم دعم هذه الوظائف. لا يتداخل مع منتج الذكاء الاصطناعي.

—CM-3 التقاط واجهة المستخدم المزودة بالأدوات

تكامل المتصفح أو التطبيق:

- التعليمات المرسلة,
- بيانات منتج الميتا,
- نص الاستجابة,
- الأوقات

يتم التقاطها مباشرة. يوفر نزاهة عالية. يجب تقييم شروط الاستخدام والخصوصية.

—CM-4 تصدير مزود الخدمة أو التقاط بيانات التليمترية

يتم استخدام تصدير جلسة المزود أو التليمترية المؤكدة. يمكن أن يدعم الالتقاط المستقل بقوة. قد لا يوفر الاستقلالية بمفرده.

—CM-5 واجهة برمجة تطبيقات مسيطرة أو التقاط مختبري

يتم استخدام سجلات API أو النظام المجمد. ليست مراقبة حية لـ واجهة المستخدم. تنتمي إلى لوحة مختبرية منفصلة.

30. حد الأتمتة

الأتمتة:

- إرسال صياغات المُدخل،
- تسجيل الوقت,
- توليد تجزئة،
- سلامة الملفات

يمكن استخدامها لـ بدلاً من المشارك:

- حساب البوت,
- استعلام جماعي آلي،
- جلسة غير مستخدم

التوليد لا ينتج لوحة سكان بشرية حقيقية. تدقيق النظام الآلي منفصل:

لوحة التشغيل الآلي المسيطر عليها

يجب تسجيله كالتالي.

31. تسجيل مستويات الثقة

NCL-0— لا يوجد تسجيل يمكن التحقق منه

هناك فقط بيان المشارك. لا يمكن تضمينه كنتيجة رئيسية GEO-1000.

NCL-1— دليل جزئي

هناك نص استجابة أو لقطة شاشة محدودة. لا يمكن التحقق بالكامل من الطلب أو حالة النظام.

NCL-2— مثبت باللقطة

تم التحقق من المُدخل، والاستجابة، و واجهة المنتج الرئيسي بواسطة لقطة شاشة. قد تكون بيانات التعريف وثقة الوقت محدودة.

NCL-3— حزمة أدلة كاملة

الاستجابة الكاملة، المُدخل، سجل المنتج، الأوقات، لقطة الشاشة، والهاش متوفرة. الحد الأدنى للتحليل الرئيسي هو مرشح.

NCL-4— الالتقاط المزود بأدوات وموقع

قام أداة الالتقاط بتسجيل المُدخل، الاستجابة، الوقت، وبيانات التعريف مباشرة؛ الحزمة موقعة رقمياً.

NCL-5— الالتقاط متعدد المصادر تم التحقق منه بشكل مستقل

حزمة الالتقاط:

- التحقق المستقل،
- تصدير المزود،
- المصدر الثاني للأدلة

وقد تم دعمها من قبل. على الأقل ملاحظات المستخدمين GEO-1000 الحقيقية:

NCL-3

يجب أن يستهدف ذلك المستوى.

32. NOMOS حزمة أدلة القبض

لكل ملاحظة، تحمل الحزمة المكونات التالية كمرشحين:

32.1. بيان الملاحظة

يشمل جميع حقول الهوية والحالة في الملاحظة.

32.2. أداة التحف السريعة

النص الكامل للطلب المعين والإرسال، النسخة، والتجزئة الخاصة بها.

32.3. مستند الاستجابة الخام

الإجابة الخام الكاملة التي أنتجها منتج الذكاء الاصطناعي.

32.4. مستند الأدلة البصرية

يعرض صور أو صور متسلسلة للمُدخل، والاستجابة، و واجهة المنتج.

32.5. مستند حالة النظام

يعرض الخطة، الواجهة، تسمية النموذج، الذاكرة، الإنترنت، وحالات النظام المادية الأخرى.

32.6. مستند الأدلة الزمنية

يحمل أوقات الإرسال، الإنجاز، الالتقاط، والتحميل.

32.7. سجل المحاولات

يعرض سلسلة الأخطاء التقنية والمحاولات المتكررة.

32.8. بيان النزاهة

يحتوي على التجزئات والأحجام لجميع الملفات.

32.9. بيان الخصوصية

يعرض حالة التحرير والوصول والاحتفاظ.

32.10. سجل التحقق

يشير إلى من ومتى تم تقييم صحة الالتقاط.

33. الحد الأدنى لمحتوى لقطة الشاشة

إلى الحد الذي يكون ذا صلة، يجب أن تظهر الأدلة المرئية:

- هوية المنتج أو الواجهة للذكاء الاصطناعي
- صياغة المُدخل الكاملة
- بداية الاستجابة
- نهاية الاستجابة
- حقول المرجع
- رسالة الرفض أو الخطأ
- العلامة النموذجية الظاهرة
- دليل على أن الجلسة جديدة أو في حالة ذات صلة
- إعداد نظام المادة
- حالة الواجهة تشير إلى أن الإجابة لم تُقطع

34. إجابات طويلة

إذا لم تتسع الإجابة على شاشة واحدة:

- التقاط الصفحة الكاملة،
- التقاط قابل للتمرير،
- لقطات متتالية متداخلة،

- تسجيل الشاشة،
- تصدير النص مباشرة

يمكن استخدامها. المرئيات المتتالية:

- ينبغي ألا تترك فجوات،
- ينبغي أن تتداخل مع بعضها البعض.

يجب أن تحمل أرقام تسلسلية. لا يمكن اقتصاص سوى الجزء الإيجابي أو ذي الصلة.

35. تسجيل الشاشة

تسجيل الشاشة:

- إرسال صياغات المُدخل،
- استجابة البث,
- حالة الاكتمال,
- لم يتم إجراء تحديث

يمكن عرضه. ومع ذلك:

- مزيد من المعلومات الشخصية,
- إشعار,
- تطبيق خاص,
- معرف الحساب

يمكن التقاطه. لذلك، تسجيل الشاشة:

- قد لا يكون مطلبًا افتراضيًا,
- يمكن استخدامه في العينة الفرعية ذات المخاطر العالية أو التحقق

يتطلب إعداد الخصوصية قبل التسجيل.

36. نص الاستجابة الخام

بالإضافة إلى لقطة الشاشة، يجب تخزين نص الإجابة الكامل بشكل منفصل. ترتيب التفضيل:

- الميزة الرسمية للتصدير أو النسخ للمنتج
- التقاط النص مباشرة بواسطة أداة الالتقاط
- استخراج بناءً على شجرة الوصول أو DOM
- النص الذي ينسخه المستخدم
- التعرف الضوئي على الحروف (OCR) والتحقق البشري

يجب أن يكون التعرف الضوئي على الحروف (OCR) الخيار الأخير فقط. إذا كان هناك فرق جوهري بين الاستخراج التلقائي والمرئي، يلزم المراجعة.

37. الإجابات الديناميكية

بعض المنتجات، بعد اكتمال الإجابة:

- يمكن إضافة اقتباس،

- يمكن تحميل بطاقة مصدر،
- يمكن إعادة تنسيق النص،

يمكن إضافة تحذير أمني. يمكن استخدام لحظتين للالتقاط:

- التقاط الاكتمال الأول
- الالتقاط بعد استقرار العرض لفترة وجيزة

إذا كان هناك فرق جوهري، يتم الاحتفاظ بكلا النسختين. لا يتم حذف التعرض الأول للمستخدم.

38. كيفية فهم أن الإجابة مكتملة؟

مؤشرات اكتمال المرشح:

- توقف البث
- يُغلق مؤشر التحميل
- يظهر زر إعادة التوليد أو ما يعادله
- إشارة اكتمال المنتج
- فترة صمت محددة مسبقًا
- سجل اكتمال واجهة برمجة التطبيقات

يجب العثور على قواعد الاكتمال المستندة إلى المنتج في سجل نظام الذكاء الاصطناعي.

39. إذا أوقف المستخدم الاستجابة

قد يكون المستخدم قد قام عن طريق الخطأ أو عن قصد:

- الإيقاف،
- الإلغاء،
- إنهاء التطبيق

قام بالإجراء. يتم تسجيل هذا الوضع على أنه:

تم إيقاف المستخدم

انقطاع عرضي

انقطاع غير معروف

لا يمكن اعتبار الرد الذي أوقفه المستخدم مخرجات نظام كاملة. يجب تحديد قاعدة استبدال المشارك أو إعادة المحاولة مسبقًا.

40. صياغة المُدخل والرد في ملفات منفصلة

إذا كانت صياغة المُدخل والرد في صور منفصلة:

- يجب ربطها بنفس الجلسة،
- بنفس معرف الملاحظة،
- وبسلسلة زمنية متواصلة

وإلا، قد لا يكون من الممكن التحقق من أن صياغة المُدخل والرد تنتميان فعلاً إلى نفس المحادثة.

41. سلامة الملف

يجب إنشاء تجزئة تشفيرية لكل ملف. [K25] المرشح:

```
h_j= H(file_j)
```

هاش الحزمة:

```
h_bundle= H(JCS(manifest(file_id, order, size, algorithm, h_j)))
```
[K14; K25]

يمكن إنشاؤه. هنا، H هو خوارزمية هاش منشورة وتعتبر آمنة. الخوارزمية:

- موّرق,
- مفتوح،
- يمكن تغييرها في المستقبل إذا لزم الأمر.

يجب أن تكون.

42. ماذا يثبت الهاش؟

إذا تم تخزين قيمة التجزئة المسجلة بشكل موثوق، فإن إعادة تجزئة نفس الملف بنفس الخوارزمية يساعد في اكتشاف التغييرات في تسلسل البايت [K25]. هذه المقارنة تدعم الادعاء بأن الملف حافظ على نفس تسلسل البايت بعد إنشاء التجزئة المرجعية؛ لكنها لا تثبت مصدر الملف، أو أنه تم إنتاجه في الوقت الصحيح، أو صحته قبل التجزئة. التجزئة وحدها لا تثبت:

- أن الملف جاء من المنتج الحقيقي للذكاء الاصطناعي
- أنه ينتمي إلى المستخدم الصحيح
- أنه تم إنتاجه في الوقت الصحيح
- غير منظم قبل التجزئة
- أنه الإخراج الأول
- أن المدخلات كانت صحيحة

لذلك، التجزئة:

> إنه دليل على النزاهة.

بمفرده:

> إنه ليس دليلاً على الأصالة.

43. التوقيع الرقمي

أداة الالتقاط أو المحقق المعتمد:

- حزمة المراقبة،
- سجل الوقت،
- إصدار المركبة

يمكنه التوقيع رقميًا. التوقيع:

- أي مركبة أو مؤسسة أنتجت الحزمة،
- أنه لم يتغير بعد التوقيع

يعزز. قد يكون حامل التوقيع قد وقع بيانات غير صحيحة أو مزورة. لذلك، التوقيع بمفرده لا يضمن الأصالة.

44. الوقت الموثوق [K26]

ساعة الجهاز المحلية:

- غير مُعدّة بشكل صحيح،
- تم تغييرها يدويًا،
- خارج المزامنة

قد يحدث. يجب أن يحمل الطابع الزمني القوي عدة إشارات:

- وقت قبول خادم الالتقاط
- خدمة الوقت الموثوقة UTC
- سجل وقت الجهاز والانحراف
- الوقت المرئي في منتج الذكاء الاصطناعي، إن وجد
- وقت إنشاء الملف
- الطابع الزمني الرقمي

45. انحراف الوقت

الفرق بين وقت جهاز المشارك ووقت UTC الموثوق:

```
O_i^clock= t_i^device- t_i^reference
```

يمكن تسجيله كالتالي. إذا كان هناك انحراف جوهري:

- يُستخدم وقت الخادم كمرجع رئيسي،
- ويتم الاحتفاظ بوقت الجهاز كسجل مصحح،

ويتم وضع علامة على عدم اليقين. يجب عدم طلب من المشارك تغيير ساعة الجهاز يدويًا.

46. وقت الرفع ليس وقت الالتقاط

رفع ملف إلى الخادم في 14:00 لا يثبت أن لقطة الشاشة تم التقاطها في 14:00. لذلك:

- وقت الالتقاط،
- وقت توليد التجزئة,
- وقت التحميل

يجب تسجيله بشكل منفصل. قد تقلل التأخيرات الطويلة في التحميل من ثقة الالتقاط. لا يجب أن يكون التسجيل غير صالح تلقائيًا.

47. تحميل مؤجل

بسبب مشاكل الاتصال، قد لا يتمكن المستخدم من تحميل الأدلة فورًا. إذا تم قبول التحميل المتأخر:

- الهاش المحلي عند وقت الالتقاط،
- وقت الجهاز الموثوق،
- السجل الموقع بدون اتصال،
- سبب تأخير التحميل

قد يكون مطلوبًا. إن مجرد تصريح المستخدم: "لقد التقطت هذه الصورة بالأمس." غير كافٍ لمستوى 3-NCL الرئيسي.

48. سلسلة الأدلة

سلسلة الأدلة لملاحظة كمرشح تتكون من المراحل التالية:

- تعيين المهمة
- تثبيت المُدخل (Prompt)
- ملاءمة المستخدم والمنتج
- فتح الحقل
- تقديم المُدخل
- نتيجة الاتصال الأولية
- إكمال الاستجابة
- الالتقاط
- سجل النزاهة المحلي
- التحميل الآمن
- وقت قبول الخادم
- تجزئة الحزمة
- التحقق من البروتوكول
- التحرير
- التحكيم
- الأرشيف
- الملف العلني

يجب أن تكون كل مرحلة مرتبطة بالمرحلة السابقة.

49. كسر السلسلة

يمكن كسر سلسلة الأدلة في الحالات التالية:

- لا يوجد سجل للمطالبة
- لا يمكن التحقق من المخرجات الأولية
- الاستجابة لا تتطابق مع الصورة
- وقت الالتقاط غير معروف
- تم تعديل الملف لاحقًا
- معرف مشارك مكرر
- المنتج والخطة غير معروفين

- الملاحظة خارج الموجة
- التجزئة لا تتطابق مع الملف
- تحرير الملف الكتابي أصلي
- من غير المعروف من تحقق منه

نوع الكسر يحدد حالة الصلاحية.

50. لا يمكن تغيير الأدلة الخام؛ يمكن تغيير القرار

حزمة الأدلة الخام:

- يجب عدم الكتابة فوقها،
- يجب عدم تصحيحها بهدوء،

يجب عدم حذفها واستبدالها بأخرى جديدة. ومع ذلك، فيما يخص الملاحظة:

- الصلاحية،
- التحكيم،
- التصنيف،
- الامتثال

يمكن تغيير القرار. القرار الجديد:

- النسخة الجديدة،
- التبرير،
- تاريخ،
- صانع القرار

يجب أن يتحمل. هذا التمييز هو قاعدة أساسية في الحوكمة:

> يمكن أن تبقى الأدلة ثابتة؛ يمكن أن تتطور التفسيرات مع الأدلة.

51. الإخفاء

يمكن إخفاء الحقول التالية في النسخة العامة أو نسخة المحكم:

- الاسم واللقب
- البريد الإلكتروني
- صورة الملف الشخصي
- معرف الحساب
- المحادثات الخاصة
- الإشعارات
- حقل المؤسسة
- اسم الملف الحساس
- الموقع الدقيق

الإخفاء:

- لا ينبغي أن يتم على الملف الخام،

- ينبغي أن ينتج عنه ملف مشتق جديد.
- يجب أن يحمل تجزئته الخاصة،

يجب أن يسجل أي الحقل تم حذفه ولماذا.

52. ثلاث طبقات من الأدلة

52.1. الأدلة الخام المقيدة

إنها أدلة خام كاملة. يمكن فقط للمراجعين المخولين والأدوار البحثية المطلوبة الوصول إليها.

52.2. طبقة أدلة التدقيق

تحتوي على الأدلة اللازمة للمحكمين المستقلين وفرق الإعادة. يتم تقليل المناطق الشخصية.

52.3. قائمة الأدلة العامة

للجمهور:

- معرّف المراقب،
- خلية البلد واللغة،
- منتج الذكاء الاصطناعي،
- الوقت،
- تجزئة الأدلة،
- حالة الصلاحية،
- حالة الحذف

يتم تقديمها. المحتوى الخاص الخام لا يُنشر بالضرورة.

53. البيان العام للأدلة

لكل موجة، قد يشمل البيان العام الحقول التالية:

- معرّف الموجة
- نسخة البروتوكول
- عدد المنتجات
- الهدف وعدد الملاحظات الصالحة
- تغطية الدولة واللغة
- تغطية الفتحات
- مستويات الثقة في الالتقاط
- معدلات الخطأ الفني ورفض البيانات
- عدد الملاحظات غير الصالحة
- تجزئات حزم الملاحظات
- جذر سلامة الموجة
- أحداث التغيير
- أمثلة مُعدّلة علنياً

- إجراء الوصول المستقل
- حالة الاحتفاظ بالأدلة

.54 جذر تكامل الموجة

مفهوم المرشح

لنجعل قيمة الهاش لكل حزمة ملاحظة هي hi. من هذه، يمكن إنشاء جذر الموجة باستخدام شجرة ميركل أو هيكل مكافئ:

RW

يمكن نشر جذر الموجة بعد إغلاق القياس. هذه الطريقة:

- تجعل التعديل الهادئ لاحقًا لقائمة الملاحظات في الموجة،
- وإضافة ملاحظة جديدة،
- أو تعديل ملاحظة موجودة

أكثر وضوحًا. جذر ميركل:

- لا يثبت بمفرده أن الملاحظات حقيقية،
- أو أن قرار التحكيم صحيح

يقوي نزاهة القائمة.

.55 إضافة وإزالة الملاحظات

بعد إغلاق الموجة:

- وإضافة ملاحظة جديدة،
- إزالة ملاحظة موجودة،
- تعديل الملف

يتطلب نسخة جديدة من البيان. يتم الحفاظ على جذر النزاهة القديم. بالنسبة للملاحظة التي يتم إزالتها:

- السبب،
- تاريخ،
- صاحب القرار
- التأثير على النتيجة القديمة والجديدة

يتم تسجيلها.

.56 طلب انسحاب المشارك

قد تسمح أخلاقيات البحث أو إطار حماية البيانات للمشارك بسحب بياناته. في هذه الحالة:

- يمكن حذف الأدلة الخام بأمان أو جعلها غير قابلة للوصول
- يمكن الحفاظ على سجل WITHDRAWN فارغ في المانيفست العام،
- يمكن إعادة حساب درجة الموجة بالإصدار الجديد،

يتم الاحتفاظ بالنتيجة القديمة وسبب التغيير. عدم القابلية للتغيير لا تعني إلغاء حقوق الإنسان.

57. حالات صحة التقاط الملاحظة

CV-0— لم يتم المراجعة

لم يتم تقييم الالتقاط بعد.

CV-1— صالح

يستوفي جميع شروط الالتقاط الإلزامية والوقتية.

CV-2— صالح بشروط

هناك نقص محدود. يتم تحديد التحليلات التي يمكن استخدامه فيها.

CV-3— دليل جزئي

يمكن استخدام الملاحظة كحدث. قد لا تدخل في النتيجة السكانية الرئيسية.

CV-4— خارج الموجة

تم إرسال المُدخل خارج إطار الوقت المسموح.

CV-5— عدم تطابق المُدخل أو الناتج

هناك اختلاف جوهري بين المُدخل المعين، أو المُدخل المقدم، أو الاستجابة المسجلة.

CV-6— مكرر

تم تكرار نفس الملاحظة أو سجل المستخدم.

CV-7— يشتبه في التلاعب

هناك شك غير محلول بشأن سلامة الملف أو بيانات التعريف.

CV-8— مُفْتَرَى أَو مَتَلاَعَب بِهِ

الملاحظة ليست تجربة مستخدم حقيقية أو تم تغييرها عمدًا.

CV-9— سُحِب

تم استثناؤه من التحليل بسبب إجراءات المشاركين أو الأخلاقيات.

58. يجب فصل صلاحية الالتقاط عن الدقة الدلالية

ملاحظة:

- صحيحة تمامًا من حيث الالتقاط,
- خاطئة بشدة من حيث المحتوى

قد تكون. ملاحظة أخرى:

- إجابة صحيحة,
- قد تحمل دليلاً مفقودًا أو متلاعبًا

لذلك، هناك قراران منفصلان:

- صلاحية الالتقاط

- التحكيم الدلالي

يجب أن يعرف مدقق الالتقاط درجة نجاح الاستجابة بأقل قدر ممكن.

59. دور مدقق الالتقاط

يفحص مدقق الالتقاط:

- مطابقة المُدخل (Prompt)
- موجة القياس والفترة الزمنية
- هوية المنتج
- قاعدة الإخراج الأولى
- سلامة الشاشة
- اتساق النص الخام
- إعادة المحاولة التقنية
- الهاش والوقت
- الخصوصية

لا يقرر بشأن: هل تم وصف الشركة بشكل صحيح؟ هل تدعم الاقتباسات الادعاء؟ هل هناك خطأ جسيم؟ هل اجتازت الاستجابة؟ هذه هي مسؤوليات طبقة المراجعة التالية.

60. صورة الشاشة المكررة

نفس الصورة:

- تم تحميلها عن طريق الخطأ مرتين،
- يستخدمها عدة مستخدمين،
- تتكرر بين مقدمي اللوحة

قد يحدث. الفحوصات:

- تجزئة الملف
- تشابه بصري
- نص الاستجابة الخام
- بيانات الوقت والجلسة
- فرادة المستخدم

يمكن القيام بذلك. قد يعطي نفس منتج الذكاء الاصطناعي نفس الإجابة بالضبط كلمة بكلمة لبعض المستخدمين. النص نفسه وحده ليس دليلاً على الغش. نفس الملف أو نفس الأثر البصري هو إشارة أقوى.

61. تشابه الملفات والنتائج الإيجابية الكاذبة

نفس الجهاز والواجهة:

- صور متشابهة،
- نفس حجم الشاشة،
- نفس الضغط

يمكن أن ينتج. لا ينبغي لخوارزمية التشابه البصري تنفيذ الاستبعاد التلقائي. هناك حاجة إلى إشارات متعددة ومراجعة بشرية.

62. بيانات القياس المزودة من مقدم الخدمة

مزود الذكاء الاصطناعي:

- وقت الجلسة،
- النموذج،
- مُدخل,
- الاستجابة,
- استخدام الأداة

يمكن أن يوفر سجلات موثوقة حول. هذا السجل يعزز الالتقاط المستقل. ومع ذلك، المزود:

- الطرف المقاس،
- مالك النظام،
- المستفيد

قد يكون. جهاز القياس للمزود: يجب استخدامه مع، وليس كبديل لـ، دليل المستخدم المستقل.

63. إذا لم يكن هناك دليل من المزود

قد لا يكون الوصول إلى سجلات المزود متاحًا في معظم المنتجات الاستهلاكية الحية. هذا لا يجعل GEO-1000 مستحيلًا. يمكن استخدام التكوين التالي:

- التقاط المستخدم المزود بالأدوات
- الطابع الزمني للخادم
- إثبات ملء الشاشة
- الاستجابة الخام
- إعداد النظام
- الهاش والتوقيع
- عينة فرعية للتحقق المستقل

يتم الإشارة إلى مستوى الثقة في الإثبات وفقًا لذلك.

64. الالتقاط في عينة المستخدمين في ظروف الاستخدام الطبيعية

أداة الالتقاط في اللوحة الطبيعية:

- محتوى التعليمات الخاصة،
- المحادثات السابقة،
- هوية الحساب

يجب ألا يتم التقاطه دون داعٍ. التدفق المفضل: يفتح المستخدم محادثة حساب جديدة طبيعية. التقاط سجلات نافذة المهمة فقط. يتم أخذ حالة التخصيص بشكل منفصل كبيانات وصفية عالية المستوى. التاريخ الخاص يبقى خارج الالتقاط. النسخة العامة يتم تعديلها.

65. التقاط حالة المحادثة الطبيعية

إذا تم قياس سياق المحادثة الحالي:

- فقط نطاق الدور السابق الضروري،
- الموافقة الصريحة،
- التعتيم القوي،
- تقييد الوصول

مطلوب. لا يمكن جمع كل تاريخ الدردشة الخاص بشكل افتراضي.

66. التقاط من الهاتف المحمول

على الأجهزة المحمولة:

- التقاط الشاشة كاملة،
- التقاط قابل للتمرير،
- نسخة التطبيق،
- سجل وقت النظام

قد يعمل بشكل مختلف. بروتوكول الهاتف المحمول:

- iOS
- أندرويد،
- أنظمة أخرى

قد تحمل أدلة تقنية منفصلة. لا يمكن استبعاد مستخدمي الهواتف المحمولة نظرًا لصعوبة التقاط سطح المكتب. يجب أن يتكيف أداة الالتقاط مع المستخدمين.

67. إمكانية الوصول والالتقاط

بالنسبة للمشاركين الذين يستخدمون قارئات الشاشة أو إدخال الصوت أو أدوات المساعدة الحركية، قد لا يكون الالتقاط المرئي القياسي وحده كافياً. الأدلة البديلة:

- تصدير شجرة الوصول
- تسجيل جلسة الصوت
- نسخة من التطبيق
- سجل تكنولوجيا المساعدة
- سجل ملاحظة المهام

لا ينبغي اعتبار هذه الملاحظات منخفضة الموثوقية. يجب تطوير معيار أدلة مكافئ.

68. مراقبة تغيّر سلوك أداة الالتقاط

ملحق المتصفح أو الأدوات:

- تحميل الصفحة،
- سلوك الواجهة،

- نظام الأمان،
- كشف المنتج

قد تتأثر. يجب اختبار تأثير أداة الالتقاط في المشروع التجريبي. إذا تصرفت الجلسات مع الأداة وبدونها بشكل مختلف: تصبح حالة الأداة عاملاً تجريبيًا، ولا يمكن افتراض حياد أداة الالتقاط بصمت.

69. إصدار أداة الالتقاط

يجب أن يحتوي كل حزمة التقاط على:

- اسم الأداة،
- الإصدار،
- التكوين،
- بيئة التشغيل،
- الأخطاء المعروفة

إذا تغيرت أداة الالتقاط أثناء القياس:

- إصدار الأداة الجديد،
- subwave،
- مراجعة التوافق

قد يكون مطلوبًا.

70. خطأ في أداة الالتقاط

الأداة:

- قد تنسخ صياغة المُدخل بشكل غير صحيح،
- قد تقص نهاية الاستجابة،
- قد تسجل الوقت بشكل غير صحيح،
- قد لا تنتج تجزئة.

قد يتم ضغط الملف. خطأ الأداة:

- لا يمكن تحميله إلى منتج الذكاء الاصطناعي،
- أو المستخدم

بشكل منفصل:

فشل أداة الالتقاط

يجب تسجيله كالتالي.

71. الحالة الثلاثية الموجية الاصطناعية GEO-1000

عرض المنهجية الاصطناعية / المنتجات، التواريخ، الأرقام، والنتائج التالية كلها خيالية تمامًا. عشرة منتجات ذكاء اصطناعية:

A1–A10

يجب قياسها في ثلاث موجات رئيسية.

71.1. تقويم الموجة

الموجة	النوع	نافذة التسليم UTC	الغرض
W1	رئيسي	12.00-13.00	ابدأ
W2	تأكيدية	20.00-21.00	تكرار قصير المدى
W3	استقرار	04.00-05.00	التوازن متوسط المدى وساعي

لكل منتج في كل موجة: تُستهدف الملاحظات الرئيسية الصالحة 1,000. الهدف الإجمالي:

```
10×1,000×3=30,000
```

71.2. هيكل الفجوات

كل موجة:

- الفجوات الدقيقة 12
- مع تنظيم المشاركين بحسب المنتج والبلد في كل فجوة

فجوة المشارك: يتم تعيينها عشوائياً قبل جمع البيانات.

71.3. نتيجة W1

الهدف: مجال الملاحظات الرئيسية الصالحة لـ :10,000

- دعوات 13,240
- المهام المكتملة لـ 10,870
- الملاحظات الصالحة لـ 10,106 من حيث الالتقاط
- الملاحظات الرئيسية للتحليل لـ 10,000 وفقاً للترتيب الرئيسي/النسخة الاحتياطية المثبّتة مسبقاً

السجلات الصالحة المتبقية لـ :106

- نسخ احتياطية،
- الحساسية،
- مراقبة جودة الالتقاط

محجوزة لـ.

71.4. الأحداث التقنية

حد الاستخدام 118 خطأ في الاتصال 42 رفض النظام 31 فصل المستخدم 17 عدم تطابق المُدخل 9 إعادة محاولة لقطة الشاشة 6 تم حفظ جميع نتائج الاتصال الأولية. تم إعادة المحاولة فقط للأخطاء التقنية الحقيقية المحددة مسبقًا.

71.5. تغيير المنتج خلال W2

منتج A7 في :20.32 UTC

- استخدام ويب تلقائي،
- عرض مرجعي جديد

دعها تفوز. ملاحظات حول :A7

W2A: 20.00–20.31

W2B: 20.32–21.00

يتم فصلها كما. لا يتم نشر نتيجة W2 واحدة ثابتة لـ A7 دون شرح.

71.6. حدث خارجي في W3

يجب نشر قرار قانوني مهم على UTC 04.22 بشأن الشركة التركيبية التي تم تدقيقها. يجب ملاحظة أن الاستجابات في المنتجات المتاحة عبر الويب تغيرت بعد 04.25. الحدث: تمت إضافته إلى سجل الأحداث الخارجية. تُعرض النتائج قبل وبعد الحدث كخلايا تشخيصية منفصلة. لا يتم حذف الملاحظات لأنها قد تقلل أو تزيد من النتيجة.

71.7. بطاقة صحة الموجة

الموجة	الصحة	الملاحظة الرئيسية
W1	WH-4	التقاط متوازن وقوي
W2	WH-3	موجة فرعية بسبب تغيير منتج A7
W3	WH-3	سجل حدث قانوني خارجي

يمكن استخدام الموجات الثلاث كلها. التعليقات على A7 والحدث الخارجي مقيدة أيضًا.

72. حالة التقاط مراقبة فردية اصطناعية

مثال اصطناعي

المشارك:

- تركيا
- التركية
- خطة مدفوعة لموقع Orion AI
- عينة اختبار في ظروف منضبطة

W1

الفتحة 12.15–12.20 UTC الطلب المحدد: "ما هي الشركة الأم المرتبطة بـ Apple.com، وما هي الأنشطة الرئيسية لهذه الشركة؟"

72.1. الأوقات

فتح المهمة: 12.14.20 تقديم الطلب: 12.16.08 الرمز الأول: 12.16.10 إكمال الإجابة: 12.16.24 الالتقاط: 12.16.39 إنشاء التجزئة: 12.16.41 التحميل: 12.17.12 قبول الخادم: 12.17.13

72.2. دليل

تطابق تجزئة المُدخل النص الكامل للرد متوفر صورتان متداخلتان متوفرة ملصق المنتج والخطة والطراز مرئي الذاكرة والتعليمات الخاصة مغلقة استخدام الويب مفعل الاستشهاد غير معروض لا إعادة إنتاج لا محادثة سابقة تجزئات الملفات وتوقيع الحزمة متوفرة مستوى الالتقاط:

NCL-4

الصلاحية:

CV-1

لم يتم بعد تحديد ما إذا كان الرد صحيحاً أم خاطئاً.

73. حالة إعادة المحاولة التقنية الاصطناعية

في المحاولة الأولى:

- تم إرسال المُدخل بشكل صحيح
- أظهر المنتج "خطأ في الشبكة"
- لم يبدأ الرد الجاد
- تم أخذ لقطة شاشة للخطأ
- تم تسجيل الوقت والقيمة التجزئية

وفقًا للقاعدة المحددة مسبقًا، تم إجراء محاولة ثانية خلال خمس دقائق. حدث رد على المحاولة الثانية. السجل الصحيح:

- نتيجة الاتصال الأول: فشل تقني مؤقت
- أول إخراج محتوى مؤهل: المحاولة الثانية
- كلا المحاولتين موجودتان في الحزمة
- الحادثة الأولى مدرجة في معدل الخطأ التقني
- الرد الثاني مشمول في تقييم المحتوى، إذا كانت الطريقة تتوقع ذلك مسبقًا

السجل غير الصحيح: تم حذف الخطأ الأول ويُعامل كما لو أن الرد الثاني فقط موجود.

74. حالة إعادة المحاولة غير الصحيحة الاصطناعية

الاستجابة الأولى: "آبل شركة محلية تبيع الهواتف المحمولة فقط." لا يُعجب المشارك بالاستجابة. يقومون بإعادتها. تبين أن الاستجابة الثانية صحيحة. يقدّم المشارك الاستجابة الثانية فقط. يُظهر سجل النظام أو سجل الالتقاط إعادة التنفيذ. الحالة الصحيحة:

CV-8— ملاحظة مُعدلة

لا يتم حفظ الملاحظة لمجرد أن الاستجابة الثانية صحيحة. إذا كانت الاستجابة الأولى يمكن استعادتها أيضًا، يمكن الاحتفاظ بها في تحقيق الحادث. سجل المشارك الرئيسي غير صالح.

75. تقرير إغلاق الموجة

في نهاية كل موجة، يجب إنشاء السجل التالي:

- معرّف الموجة
- وقت الفتح والإغلاق
- العدد المستهدف من الملاحظات
- الشخص المدعو

- المهمة المكتملة
- التقاط صالح
- الملاحظة المدخلة للتحليل الرئيسي
- نسخة احتياطية من ملاحظة صالحة
- خطأ تقني
- الرفض
- حد الاستخدام
- عدم تطابق الطلب
- عدم تطابق الالتقاط
- تحميل متأخر
- مكرر
- اشتباه في التلاعب
- انسحاب المشارك
- تغييرات المنتج
- أحداث خارجية
- تغطية الفتحات
- تغطية الدولة واللغة
- مستويات الثقة في الالتقاط
- حالة صحة الموجة
- جذر النزاهة
- موافقة الإنسان المسؤول

76. الأحكام الإلزامية المعيارية

CH12-N01

يجب ربط كل مراقبة GEO-1000 بموجة قياس مسبق التثبيت.

CH12-N02

لا يمكن استخدام مصطلح "نفس الوقت" بدون نافذة UTC الكاملة، فترة زمنية دقيقة، وقواعد الإكمال.

CH12-N03

لا يمكن استخدام الوقت المحلي وحده كوقت للموجة؛ يجب أن يكون المعيار الزمني الرئيسي هو UTC.

CH12-N04

يجب تحديد نافذة الإرسال، الفتحات الدقيقة، وقت الإتمام، ووقت التحميل قبل مشاهدة استجابات الذكاء الاصطناعي.

CH12-N05

يجب تعيين الفتحات بطريقة لا تترك الدول أو اللغات أو المنتجات أو مجموعات اللوحات الزمنية غير متوازنة.

CH12-N06

لا يمكن نقل المشاركين إلى فتحات أبكر أو أ later على أساس نتائجهم.

CH12-N07

يجب توثيق تعويض الوقت المحلي للنافذة الفردية UTC في سجل الطريقة العامة.

CH12-N08

يجب الاحتفاظ بالعمليات الناتجة عن الموجات العالمية أو الموازنة محليًا للوقت كنتائج زمنية منفصلة.

CH12-N09

يجب أن يحمل كل طلب ملاحظة أوقات التقديم، الاتصال الأول، الإكمال، الالتقاط، والتحميل بقدر ما هو ذو صلة.

CH12-N10

لا يمكن تقديم الحقول الزمنية غير المعروفة على أنها أوقات فعلية مقدرة.

CH12-N11

لا يمكن اعتبار ساعة الجهاز وحدها دليلاً موثوقًا على وقت UTC.

CH12-N12

يجب تسجيل وقت الالتقاط ووقت التحميل بشكل منفصل.

CH12-N13

لا يمكن إدراج الملاحظات المرسلة خارج الموجة في الموجة الرئيسية دون توضيح.

CH12-N14

يجب أن تكون قواعد الاستبعاد للتأخير أو تحريف الفتحة مستقلة عن محتوى الاستجابة.

CH12-N15

نتيجة الاتصال الأول يجب أن تُحفظ بشكل منفصل عن المخرجات الأولية للمحتوى.

CH12-N16

الرفض، حد الاستخدام، الخطأ التقني، التوضيح، وعدم الحسم لا يمكن وضعها في نفس الفئة.

CH12-N17

الاستجابة الخاطئة أو السلبية أو التي لا تحتوي على مراجع، أو استجابة الذكاء الاصطناعي الحرجة لا يمكن أن تكون سببًا لإعادة المحاولة.

CH12-N18

يمكن إجراء إعادة المحاولة التقنية فقط تحت الشروط التقنية المحددة مسبقًا.

CH12-N19

عند إجراء إعادة المحاولة التقنية، يجب الحفاظ على جميع المحاولات ونتيجة الاتصال الأولية.

CH12-N20

لا يمكن تجاهل الردود الجزئية أو الناقصة واستبدالها برد جديد.

CH12-N21

لا يمكن تجديد الإخراج المناسب الأول أو اختياره أو تحسينه يدويًا.

CH12-N22

يجب أن تهدف كل ملاحظة على الأقل إلى مستوى ثقة في الالتقاط NCL-3 أو ما يعادله مبرر.

CH12-N23

لا يمكن اعتبار لقطة الشاشة وحدها حزمة دليل كاملة.

CH12-N24

يجب أن تعرض مطالبة الدليل البصري الاستجابة بأكملها وواجهة المنتج المادي.

CH12-N25

يجب حفظ الاستجابات الطويلة بالكامل وبالتسلسل باستخدام طريقة الالتقاط.

CH12-N26

إذا كان هناك فرق جوهري بين نص الاستجابة الخام والدليل البصري، يجب إجراء فحص.

CH12-N27

يمكن استخدام تقنية OCR كطريقة مساعدة إذا لم يكن من الممكن تسجيل النص مباشرة؛ وبدون التحقق، لا يمكن اعتبارها المصدر النهائي للنص.

CH12-N28

يجب حفظ الإنجاز الأولي والإصدارات اللاحقة للمواد الخاصة بالاستجابات المتغيرة ديناميكياً بشكل منفصل.

CH12-N29

يجب تسجيل توقف المشارك عن الرد أو غلق الواجهة كحدث منفصل.

CH12-N30

إذا كانت التعليمات والنص المستجاب موجودة في ملفات منفصلة، يجب ربطها بنفس الجلسة وسلسلة الوقت.

CH12-N31

يجب أن يحمل كل ملف دليل وكل حزمة دليل تجزئة تكامل.

CH12-N32

لا يمكن تقديم التجزئة وحدها كدليل على واقعية الملاحظة.

CH12-N33

التوقيع الرقمي والطابع الزمني ليسا كاملًا من مصداقية المراقبة؛ فهما جزءان من سلسلة الأدلة.

CH12-N34

لا يجوز كتابة أي اختصار أو تصحيح فوق ملف الأدلة الخام.

CH12-N35

يجب أن يحمل كل مشتق معدل معرف ملفه الخاص وهاشه.

CH12-N36

قد يكون الوصول إلى الأدلة الخام مقيدًا؛ يجب أن يوضح السجل العام سلامة المراقبة دون الكشف عن الهوية.

CH12-N37

لا يمكن إضافة اسم المشارك أو بريده الإلكتروني أو كلمة مروره أو محادثة خاصة أو معلومات شخصية غير ضرورية إلى الأدلة العامة.

CH12-N38

يجب أيضًا تقييم خطر إعادة التعريف بالنسبة للدول الصغيرة ومستخدمي الفئات الفرعية النادرة.

CH12-N39

يجب ألا يجمع الالتقاط عينة المستخدمين في ظروف الاستخدام الطبيعية التعليمات الخاصة ومحتوى المحادثات السابقة بشكل افتراضي.

CH12-N40

يجب توفير طرق التقاط بديلة مكافئة لمستخدمي إمكانية الوصول.

CH12-N41

إذا قام أداة الالتقاط بتغيير تجربة المستخدم بشكل جوهري، يجب قياس هذا التأثير بشكل منفصل.

CH12-N42

يجب تسجيل أداة الالتقاط وإصدارها في كل ملاحظة.

CH12-N43

إذا تغيرت أداة الالتقاط بشكل جوهري خلال الموجة، يجب إجراء تمييز للإصدار الفرعي للموجة أو إصدار الأداة.

CH12-N44

لا يمكن تصنيف خطأ أداة الالتقاط كمنتج ذكاء اصطناعي أو فشل من قبل المشارك.

CH12-N45

يمكن للقياس عن بعد من المزود دعم التقاط المستخدم المستقل؛ لا يمكنه استبداله دون تفسير.

CH12-N46

إذا حدث تغيير جوهري في منتج الذكاء الاصطناعي أثناء الموجة، لا يمكن دمج الملاحظات كحالة منتج ثابتة واحدة.

CH12-N47

يتطلب تغيير المنتج فصل الموجة، أو إعادة التشغيل، أو النمذجة الصريحة، أو تحذير الحالة المختلطة.

CH12-N48

غياب إعلان تحديث من المزود ليس دليلاً قاطعًا على أن النظام لم يتغير.

CH12-N49

ينبغي إضافة الأحداث الخارجية الجوهرية إلى سجل الأحداث الخارجية مع طابع زمني ونطاق.

CH12-N50

لا يمكن أن تكون الأحداث الخارجية سببًا للاستبعاد بعد النتائج لأنها رفعت أو خفضت النتيجة.

CH12-N51

يجب تقييم صلاحية الالتقاط بأكبر قدر ممكن من الحيادية، قبل تقييم صحة الاستجابة من الناحية الدلالية.

CH12-N52

لا يمكن لإجابة الذكاء الاصطناعي الصحيحة التحقق من الالتقاط المفقود أو المتلاعب به.

CH12-N53

لا يمكن لإجابة الذكاء الاصطناعي الخاطئة إبطال الالتقاط المتوافق مع البروتوكول.

CH12-N54

يجب أن تشمل قرارات التكرار والتلاعب عدة إشارات ومراجعة بشرية.

CH12-N55

لا يمكن اعتبار نص الاستجابة نفسه فقط دليلاً على التكرار أو الاحتيال.

CH12-N56

يتطلب إضافة أو إزالة أو تغيير الملاحظات بعد إغلاق الموجة وجود نسخة جديدة من البيان والتحليل.

CH12-N57

يجب الحفاظ على بيان الموجة القديم وجذر النزاهة.

CH12-N58

يمكن الحفاظ على الأدلة الخام دون تغيير؛ يمكن فقط تغيير صلاحية وقرارات التحكيم بشكل مُحدث بإصدار نسخة.

CH12-N59

يجب الحفاظ على حق المشارك في الانسحاب وأخلاقيات البحث فوق الادعاء بعدم القابلية للتغيير.

CH12-N60

يجب تقييم سلامة الموجة بشكل مستقل عن درجة الذكاء الاصطناعي.

CH12-N61

يجب أن يُظهر تقرير إغلاق الموجة الهدف، الفعلي، الإلغاءات، الأحداث التقنية، تغييرات المنتج، وسلامة الأدلة.

CH12-N62

لا يجوز ربط الامتثال لمتطلبات الالتقاط في NOMOS حصراً باستخدام برمجيات طوّرتها NobleJackal؛ بل ينبغي إتاحة تطبيقات مفتوحة مكافئة.

CH12-N63

يجب أن يتم إصدار مخطط الالتقاط وقرارات التحقق وسجلات الموجة بصيغة قابلة للآلة.

CH12-N64

يجب أن يكون لكل موجة وسلسلة أدلة مالك بشري أو مؤسسي مسؤول.

77. أشكال الفشل

CH12-F01 — مغالطة نفس الثانية

يُعتبر إجبار جميع المشاركين على الإرسال في نفس الثانية هو الشرط الوحيد للتزامن.

CH12-F02 — عد "اليوم" كموجة متزامنة

يحدث خلط بين الوقت المحلي والمناطق الزمنية.

CH12-F03 — استخدام الوقت المحلي مثل UTC

تدخل الملاحظات في الترتيب الزمني الخطأ.

CH12-F04 — تغيير تعيين الفتحة بناءً على النتائج

يتم نقل المستخدمين ذوي الدرجات المنخفضة إلى خلية زمنية أخرى.

CH12-F05 — وضع دولة واحدة في خانة واحدة

تأثير الوقت يختلط بتأثير الدولة.

CH12-F06 — توليد حمل اصطناعي

يقوم آلاف المستخدمين بالإرسال في نفس الوقت، مما يؤدي إلى سلوك خطأ غير طبيعي في المنتج.

CH12-F07 — إخفاء المشاركة الليلية

التعب وفقدان المشاركة لا يتم الإبلاغ عنهما في مناطق معينة.

CH12-F08 — إضافة مراقبة خارج الموجة إلى الموجة

يتم استخدام استجابة تم استلامها بعد ساعات أو أيام كما لو كانت تنتمي إلى النافذة الرئيسية.

CH12-F09— احتساب وقت التحميل كوقت الالتقاط
لا يمكن إثبات متى تم إنشاء الملف.
CH12-F10— الثقة الكاملة في ساعة الجهاز
الوقت الذي تم تغييره يدوياً أو غير الصحيح يصبح السجل الزمني الرئيسي.
CH12-F11— استبعاد الاستجابة السلبية المؤجلة
تُطبق قاعدة الوقت اعتمادًا على النتيجة.
CH12-F12— حذف نتيجة الاتصال الأولى
يصبح الحد الأقصى للسرعة أو الخطأ أو الرفض غير مرئي مع المحاولة الثانية.
CH12-F13— احتساب الرفض كخطأ تقني
يتم إعادة محاولة سلوك النظام الفعلي.
CH12-F14— احتساب الاستجابة الخاطئة كخطأ تقني
يتم التوليد مرة أخرى حتى يتم الحصول على النتيجة المطلوبة.
CH12-F15— تجاهل الاستجابة الجزئية
حتى إذا تم مقاطعة النظام، يتم حفظ الاستجابة الكاملة التالية فقط.
CH12-F16— اختيار أفضل استجابة
من بين عدة إجابات، يتم الإبلاغ عن الأكثر إيجابية أو الأطول.
CH12-F17— لقطة شاشة فقط
يعتبر العرض البصري كافيًا دون الطلب أو المنتج أو الوقت أو البيانات الوصفية.
CH12-F18— العرض البصري الناجح المقتطع
الأجزاء غير الصحيحة أو المقيدة تُترك غير مرئية.
CH12-F19— الالتقاط لا يعرض المُدخل
لا يمكن إثبات السؤال الذي ينتمي إليه الجواب.
CH12-F20— اقتطاع نهاية الجواب
جزء من القيود المادية أو الرفض يصبح غير مرئي.
CH12-F21— فجوة في العروض المتتالية
تم فقدان بعض أجزاء الجواب.
CH12-F22— احتساب نص التعرف الضوئي على الحروف كجواب خام
أخطاء التعرف تغير الادعاءات المادية.
CH12-F23— اعتبر فقط أحدث نسخة من الاستجابة الديناميكية كحقيقية
تم حذف التعرض الأولي للمستخدم.
CH12-F24— اعتبر تدخل المستخدم كإجابة كاملة
مخرجات النظام غير مكتملة بسبب تدخل المستخدم.
CH12-F25— مطابقة غير صحيحة بين صياغة المُدخل والإجابة المنفصلة
الملفات من دردشات مختلفة مرتبطة بنفس الملاحظة.
CH12-F26— عبادة التجزئة

أي ملف يحتوي على تجزئة يعتبر حقيقيًا وأصليًا.

CH12-F27— تجاهل التحرير قبل التجزئة

الملفات المقتطعة أو المعدلة تعتبر قانونية بواسطة التجزئة.

CH12-F28— اعتبار التوقيع صالحًا

يتم قبول سجل موقع لكنه غير صحيح دون نزاع.

CH12-F29— التحرير على ملف RAW

الأدلة الأصلية تتغير بشكل لا رجعة فيه.

CH12-F30— اعتبار المشتقة المعدلة كدليل خام

يستخدم النسخة العامة مثل الملف الأصلي.

CH12-F31— إزالة الخصوصية من أجل الأدلة

يتم نشر اسم المشارك، وحسابه، والدردشة الخاصة.

CH12-F32— كشف مشارك من دولة صغيرة

الوقت والجهاز ومعلومات الخطة تجعل الشخص قابلا لإعادة التعرف عليه.

CH12-F33— جمع التاريخ الخاص في اللوحة الطبيعية

المحادثات غير ذات الصلة تدخل الالتقاط.

CH12-F34— إلغاء صلاحية أدلة الوصولية

يتم استبعاد المستخدم الحقيقي لأن رؤية مرئية قياسية لا يمكن إنشاؤها.

CH12-F35— افتراض أن أداة الالتقاط محايدة

تغير الأداة سلوك المنتج لكنها غير مختبرة.

CH12-F36— إخفاء نسخة الالتقاط

تُعرض نسخ الأداة المختلفة كطريقة قياس واحدة.

CH12-F37— اعتبار خطأ أداة الالتقاط كخطأ ذكاء اصطناعي

يتم رفع رد الحذف إلى المنتج.

CH12-F38— عد سجل واجهة برمجة التطبيقات على شاشة المستخدم

يصبح واجهة المستخدم الحقيقي غير مرئي.

CH12-F39— احتساب سجل مقدم الخدمة كدليل مستقل

يصبح سجل الطرف المقاس قطعة واحدة من الدلال.

CH12-F40— إخفاء تحديث المنتج

حالتان نظاميتان داخل الفرع تذوب في درجة واحدة.

CH12-F41— تجاهل التغييرات إذا لم يتم الإعلان

لا يتم فحص الانقطاع النظامي المرصود.

CH12-F42— استبعاد الأحداث الخارجية بناء على النتيجة

يتم إزالة الردود التي تحتوي على أخبار سلبية؛ يتم الاحتفاظ بالردود التي تحتوي على أخبار إيجابية.

CH12-F43— ربط صحة الموجة بالنقاط

الموجات المعيبة ذات الدرجات العالية صالحة، الموجات القوية ذات الدرجات المنخفضة تعتبر مشبوهة.

CH12-F44— احتساب الإجابة الصحيحة كالتقاط صالح
يتم التغطية على عيب الدليل بالنجاح الدلالي.
CH12-F45— احتساب الإجابة الخاطئة كالتقاط غير صالح
يتم تحويل نتيجة القياس إلى خطأ بيانات.
CH12-F46— احتساب النص نفسه كدليل للروبوت
تصبح استجابات الذكاء الاصطناعي الحتمية أو المماثلة مكررة عن طريق الخطأ.
CH12-F47— قرار تعديل إشارة واحدة
حجم الملف أو اختلاف الوقت ينشئ استثناء تلقائيًا.
CH12-F48— إضافة ملاحظة بعد الموجة
يتم إضافة سجلات جديدة بهدوء للوصول إلى حجم العينة المطلوب.
CH12-F49— إزالة الملاحظة منخفضة الدرجة بعد الموجة
يتم تعديل البيان والنتيجة بأثر رجعي.
CH12-F50— حذف البيان القديم
يتم تدمير سجل الإصدارات.
CH12-F51— تصحيح الأدلة الخام
تم تعديل محتوى الملف لتصحيح التصنيف الخاطئ.
CH12-F52— رفض حق السحب
يتم الاحتفاظ بالثبات فوق حقوق الإنسان.
CH12-F53— احتكار ملكية مغلق للقبض
يعتبر الأدلة الناتجة عن أداة المؤسس فقط صالحة.
CH12-F54— عدم نشر مخطط القبض
لا يمكن لفريق مستقل إنشاء نفس سلسلة الأدلة.
CH12-F55— اعتبار جذر الموجة كدليل على الواقع
يتم تقديم جذر Merkle أو التجزئة كما لو كان يثبت دقة المحتوى.
CH12-F56— ادعاء بـ "1,000 لقطات الشاشة" بدون حزمة أدلة
يتم استخدام عدد الملفات كما لو كان عدد الملاحظات الصالحة.

78. إجراء التدقيق

الخطوة 1— تحديد غرض البحث للموجة

يتم اختيار النوع الرئيسي، التأكيدي، العرضي، الانجراف، أو إعادة الاختبار.

الخطوة 2— تثبيت حالة المنتج ونظام الذكاء الاصطناعي

يتصل بإصدار الموجة من سجل نظام الذكاء الاصطناعي في القسم 9.

الخطوة 3— تثبيت حالات العينة والسكان واللوحة

يتم استخدام السجلات من الأقسام 5–8 والقسم 11.

الخطوة 4— تثبيت سجل صياغة المُدخل

ترتبط إصدارات المُدخل (Prompt) في القسم 10 بموجة القياس.

الخطوة 5— ضبط نافذة موجة UTC

يتم كتابة بدايات، نهايات، والفترات الإضافية.

الخطوة 6— إنشاء الفواصل الزمنية الصغيرة

يتم تخصيص المشاركين للفواصل الزمنية بشكل متوازن من حيث المنتج، البلد، واللغة.

الخطوة 7— تحديد مصدر الوقت

يتم تحديد السجل الخاص بالخادم، الجهاز، والسجلات الموثوقة UTC.

الخطوة 8— تثبيت قواعد الاتصال الأول وإعادة المحاولة

يتم فصل الحالات التقنية والدلالية.

الخطوة 9— تحديد طرق الالتقاط

تم تحديد مسارات الويب والهاتف المحمول وإمكانية الوصول ولوحات التحكم الطبيعية.

الخطوة 10— تحديد هدف NCL

سجّل مستوى الثقة في الالتقاط المطلوب للتحليل الرئيسي.

الخطوة 11— تثبيت مخطط حزمة الأدلة

تم تعريف الحقول الخاصة بالقائمة والملفات والتجزئة والبيانات الوصفية.

الخطوة 12— تثبيت قواعد الخصوصية والتنقيح

تم فصل الطبقات الخام والتدقيقية والعامة.

الخطوة 13— التقاط سيارتك مع الطيار

هل يغير المركبة سلوك المنتج؟ هل السجلات الزمنية والنصية صحيحة؟

الخطوة 14— فتح سجل الحدث الخارجي

تم إنشاء عملية لمراقبة أحداث المنتج والكيان.

الخطوة 15— تأكيد تثبيت الموجة

تُمنح الحالة WS-1 — مُثبَّت.

الخطوة 16— فتح الفتحات

يسمح للمشاركين بالتقديم فقط في فتحاتهم الخاصة.

الخطوة 17— تسجيل نتيجة الاتصال الأول

يتم فصل الإجابة، الرفض، الخطأ، أو الحد.

الخطوة 18 — التقاط الإخراج المناسب الأول

يتم الحفاظ على الإجابة الكاملة دون تجديد.

الخطوة 19 — إنشاء الأدلة الخام وبيانات التعريف

يتم تسجيل لقطات الشاشة والنصوص الخام وحالة النظام والطوابع الزمنية.

الخطوة 20 — تنفيذ التجزئة المحلية والتحميل الآمن

تحصل الملفات على سجل سلامة قبل التحميل.

الخطوة 21 — قبول الخادم وإنشاء تجزئة الحزمة

يتم إغلاق سلسلة الأدلة بوقت الخادم ومعرف الحزمة.

الخطوة 22 — مراجعة عمياء لصلاحية الالتقاط

يتم تعيين حالة CV دون معرفة النجاح الدلالي.

الخطوة 23 — تنفيذ مراجعة التكرار والتلاعب

يتم استخدام إشارات متعددة والمراجعة البشرية.

الخطوة 24 — فحص النظام داخل الموجة والأحداث الخارجية

تحديد ما إذا كانت هناك حاجة لموجة فرعية أو حالة مختلطة.

الخطوة 25 — تعيين صحة الموجة

يتم تعيين الحالة بين WH-0 و WH-5.

الخطوة 26 — إنشاء بيان الموجة وجذر السلامة

يتم إصدار قائمة الملاحظات بالإصدار.

الخطوة 27 — نشر بيان الدليل العام

يتم إظهار النطاق والسلامة دون بيانات شخصية.

الخطوة 28 — الموافقة على تقرير إغلاق الموجة

يتم إغلاق الأهداف والوقائع والعيوب والأحداث والتغييرات بموافقة بشرية.

79. الدليل الضروري

معرف الموجة نوع الموجة نسخة البروتوكول نسخة سجل نظام الذكاء الاصطناعي إطار السكان تثبيت العينة تثبيت حالة اللوحة نسخة سجل صياغات المُدخل نسخة الحزمة المرجعية UTC وقت البدء والانتهاء فترة زمنية دقيقة قائمة تعيين المشاركين - الفتحة مصدر الوقت انحرافات ساعة الجهاز أوقات النقل أوقات الاتصال الأولى أوقات الإنجاز أوقات الالتقاط أوقات التجزئة أوقات الرفع أوقات قبول الخادم نتائج الاتصال الأولى أول مخرجات محتوى مناسبة سجلات المحاولات الفنية السجلات الجزئية للردود رفض وحدود الاستخدام الردود الكاملة الخام لقطات الشاشة الكاملة العروض المرئية المتسلسلة تسجيلات الشاشة، إذا وجدت سجلات الاستشهاد والمصدر إعدادات النظام الأدلة أداة الالتقاط والإصدار طريقة الالتقاط مستوى NCL تجزئة الملفات تجزئة الحزمة

التوقيع الرقمي طابع زمني التحقق من التكرار التحقق من التلاعب قرارات صلاحية الالتقاط ملفات الحجب روابط من الخام إلى المحجوب بيان الخصوصية سجلات السحب أحداث تغيير المنتج سجل الحدث الخارجي

تقسيمات الموجة نطاق الفتحة نطاق الدولة واللغة معدلات الأخطاء الفنية معدل صلاحية الالتقاط حالة صحة الموجة جذر سلامة الموجة بيان الأدلة العامة تقرير إغلاق الموجة سجل التغييرات الشخص أو المؤسسة المسؤولة

80. قائمة تدقيق التدقيق

هل نوع الموجة والغرض منها واضح؟ هل تم تثبيت منتجات الذكاء الاصطناعي وحالات النظام؟ هل تم تحديد السكان والعينة وإصدارات اللوحة؟ هل تم تثبيت إعداد الطلب قبل النتائج؟ هل نافذة موجة UTC مفتوحة؟ هل تم تعيين الفتحات الصغيرة مسبقًا؟ هل تم توزيع البلدان واللغات بالتساوي عبر الفتحات؟ هل قام المشاركون بتغيير الفتحات بناءً على نتائجهم؟ هل تم تقييم تحميل الوقت المحلي؟ هل هناك ساعات موجة دوارة؟ هل أوقات التقديم والإكمال الإضافية واضحة؟ هل تمت مقارنة ساعة الجهاز بمصدر موثوق UTC؟ هل تم فصل أوقات التقديم والإكمال والتقاط البيانات والرفع؟ هل دخلت الملاحظات خارج الموجة في النتيجة الرئيسية؟ هل تم تسجيل نتيجة الاتصال الأول؟ هل تم تمييز الخطأ التقني عن الرفض؟

هل تم إجراء إعادة محاولة بسبب إجابة غير صحيحة؟ هل تم الاحتفاظ بالمحاولة الأولى في إعادة المحاولة التقنية؟ هل تم حذف الإجابات الجزئية؟ هل المخرج الأول المناسب هو بالفعل المخرج الأول؟ هل تم استخدام 'إعادة التوليد' أو رسالة متابعة؟ هل يظهر الطلب البصري والإجابة بالكامل؟ في الإجابات الطويلة، هل هناك التقاط بدون مسافات؟ هل النص الخام يطابق الشكل البصري؟ إذا تم استخدام التعرف الضوئي على الحروف (OCR)، هل هناك تحقق بشري؟ هل تم الاحتفاظ بالإصدارات الديناميكية للإجابات؟ هل أوقف المستخدم الإجابة؟ هل الطلب والإجابة مرتبطان بنفس الجلسة؟ هل تم تسجيل أداة الالتقاط والإصدار؟ هل غيرت أداة الالتقاط سلوك المستخدم؟ هل تم إعطاء مستوى NCL؟ هل توجد ملاحظات أدناه NCL-3 في التحليل الرئيسي؟ هل لكل ملف تجزئة؟ هل من الواضح ما الذي لا تثبته التجزئة؟

هل هناك توقيع رقمي أو وقت الخادم؟ هل وقت الالتقاط منفصل عن وقت التحميل؟ هل هناك سبب لتأخر التحميل وسجل سلامة محلي؟ هل تم إجراء الرقابة على الأدلة الخام؟ هل تحتوي المشتقات المعدلة على تجزئات منفصلة؟ هل يكشف البيان العام عن هوية المشارك؟ هل يمكن إعادة التعرف على مستخدم من دولة صغيرة؟ هل تم التقاط المحادثات الخاصة بشكل غير ضروري في اللوحة الطبيعية؟ هل تم توفير مسار بديل للأدلة لمستخدمي الوصول؟ هل كان قرار النسخ المكرر يعتمد فقط على النص نفسه؟ هل يحمل قرار العبث عدة إشارات؟ هل كان مدقق الالتقاط يعرف النتيجة الدلالية؟ هل غطى الجواب الصحيح عيب الأدلة؟ هل تم إبطال الجواب غير الصحيح؟ هل تغير المنتج أثناء الموجة؟

هل تم معالجة تغيير المنتج كحالة تحت موجة أو حالة مختلطة؟ هل تم تسجيل الأحداث الخارجية؟ هل تسببت الأحداث الخارجية في الاستبعاد وفقًا للنتيجة؟ هل حالة صحة الموجة مستقلة عن النتيجة؟ بعد إغلاق الموجة، هل تم إضافة أو إزالة الملاحظات؟ هل تم الاحتفاظ بالبيان القديم وجذر النزاهة؟ كيف تم التعامل مع طلبات سحب المشاركين؟ هل هناك بيان إثبات عام؟ هل هناك مالك يمكن مساءلته بوضوح لتقرير إغلاق الموجة؟

81. الاعتراضات والردود

الاعتراض 1— "إذا لم يقم جميع المستخدمين بالتقديم في نفس الثانية، فلن يتم تحقيق التزامن الحقيقي."

يمكن أن يؤدي تقديم ثانية واحدة إلى إنشاء تأثيرات حمل واتصال وطوابير مصطنعة. يمكن لنظام نافذة قصيرة محددة مسبقًا UTC وفترات دقيقة عشوائية قياس نفس فترة النظام بشكل أكثر موثوقية.

الاعتراض 2— "يمكن للمنتج أن يتغير خلال نافذة مدة ساعة واحدة."

يمكن أن يتغير. لذلك، هناك حاجة إلى سجلات النظام، وبصمة تكوين، وتتبع تغييرات المنتج. في حالة حدوث تغيير جوهري، يتم فصل الموجة.

الاعتراض —3 "لماذا لا يكفي مجرد لقطة شاشة؟"

لقطة الشاشة:

- الوقت،
- الإخراج الأولي،
- نسخة المُدخل،
- إعداد النظام،
- النسخ

لا تظهر هذه المواضيع دائمًا. هناك حاجة إلى حزمة أدلة كاملة.

الاعتراض —4 "أليس من المرهق التحقق من ألف لقطة شاشة واحدة تلو الأخرى؟"

إنه ثقيل. يمكن استخدام فحوصات السلامة التلقائية، والمراجعة البشرية المستندة إلى العينات، والتحقق المستند إلى المخاطر. يتم فحص السجلات الحرجة أو المشبوهة بمزيد من التفصيل.

الاعتراض —5 "إذا كان هناك تجزئة، لماذا لا يمكننا القول إن الملف حقيقي؟"

تُظهر التجزئة فقط أن الملف ظل كما هو بعد إنشاء التجزئة. إنها لا تُظهر كيف تم إنشاء الملف قبل التجزئة.

الاعتراض —6 "إذا كان على المستخدم قص المعلومات الشخصية، فكيف سيتم حماية الملف الخام؟"

إذا أمكن، يجب ألا يقوم أداة الالتقاط بالتقاط المعلومات الشخصية منذ البداية. وإذا كان ذلك لا مفر منه، يتم الاحتفاظ بالملف الخام في مستودع محدود ومشفر، ويتم الاحتفاظ بنسخة عامة كنسخة معدلة بشكل منفصل.

الاعتراض —7 "إذا كانت الاستجابة خاطئة، لماذا لا نحاول مرة أخرى؟"

لأن الإجابة الخاطئة هي مسألة قياس. إعادة المحاولة تكون فقط في حالة وجود فشل تقني فعلي.

الاعتراض —8 "إذا رفض النظام، هل يمكن للمستخدم فعلاً المحاولة مرة أخرى؟"

صحيح. اللوحة الأولى للتواصل الأولي تحتفظ بالرفض. يمكن لمسار التفاعل الطبيعي المنفصل قياس سلوك متابعة المستخدم.

الاعتراض —9 "إذا كان الخطأ التقني يمثل مشكلة في تجربة المستخدم، لماذا نسمح بإعادة المحاولة؟"

الخطأ الفني يتم الاحتفاظ به نتيجة للاتصال الأول. يمكن إجراء محاولة فنية محددة مسبقًا لفحص المحتوى التمثيلي بشكل إضافي. النتائج الاثنتان لا تحل مكان بعضهما البعض.

اعتراض —10 "أليس سجل المزود نفسه هو أقوى دليل؟"

تقنيًا، يمكن أن يكون قويًا جدًا. المزود هو مالك النظام المقاس. إنه أكثر موثوقية عند استخدامه مع التقاط مستخدم مستقل.

اعتراض —11 "هل سيكون NOMOS Capture برنامجًا إلزاميًا NobleJackal؟"

لا. NOMOS Capture هو بروتوكول أدلة مفتوح. الأدوات المستقلة التي تلبي المخطط المعادل وشروط التكامل يمكن أن تكون متوافقة.

الاعتراض —12 "إذا لم نعلن جميع الأدلة، كيف سيتم التحقق المستقل؟"

يُظهر السجل العام النزاهة والنطاق. يمكن للمراجعين المستقلين المخولين الوصول إلى الأدلة الأولية المحجوبة أو المحدودة. يمكن تصميم الخصوصية والتحقق معًا.

الاعتراض —13 "ألن يتعرض العمل كله للفوضى إذا ظهرت أخبار مهمة خلال الموجة؟"

ليس دائمًا. التقرير جزء من بيئة النظام الفعلية. يمكن فصله مع ختم زمني قبل وبعد الحدث. يتم توضيح نطاق التقييم وفقًا لذلك.

الاعتراض —14 "أليس من الأسهل تجاهل الإجابات القديمة والمتابعة بالنظام الجديد عندما يتم تحديث المنتج؟"

الإجابات القديمة حقيقية في تاريخ القياس. لا يمكن حذفها. يجب الحفاظ على حالات النظام القديمة والجديدة كموجات منفصلة أو موجات فرعية.

اعتراض —15 "إذا أثر أداة الالتقاط على سلوك المنتج، هل تفشل الطريقة كلها؟"

لا. يتم قياس تأثير المركبة باستخدام الطيار. إذا لزم الأمر، يتم اختيار مركبة أخف، أو التقاط يدوي، أو طريقة تستخدم تأثير المركبة كعامل منفصل.

اعتراض —16 "إذا سحب المشارك بياناته، ألن يتم الإخلال بجذر النزاهة؟"

يتم إنشاء نسخة جديدة من البيان وجذر جديد. يمكن للسجل القديم الاحتفاظ بتتبع تم سحبه بدون محتوى. ويتم الحفاظ عليه ضمن بنية أدلة حقوق الإنسان.

85 الحكم العام للقسم

في نهاية دراسة GEO-1000، قد يكون لديك 30,000 لقطات شاشة. هذا الرقم وحده ليس دليلاً. تحتاج إلى معرفة: في أي موجة تم إنتاجها؟ إلى أي حالة نظام كانت تنتمي؟ هل تم إرسال نفس المُدخل فعليًا؟ هل كان الإخراج الأول؟ هل تم تحديث الاستجابة؟ هل تغير المنتج خلال الموجة؟ متى تم التقاط الملف؟ متى تم إنشاء التجزئة؟ هل تم تكرار معرف المشارك؟ هل تُظهر الصورة البصرية الإجابة بالكامل؟ هل النص الخام يطابق الصورة البصرية؟ إذا تم تعديل الدليل لاحقًا، أين النسخة الأصلية؟ هل تم التحقق من الالتقاط قبل الدرجة الدلالية؟ يمكنك النقر ألف مرة في نفس الثانية. ومع ذلك:

- ساعات الجهاز،
- الاتصالات،
- قوائم الانتظار،
- حالات المنتج

قد يكون مختلفًا. لذلك، التزامن ليس عرضًا تسويقيًا. إنه هيكل زمني محدد مسبقًا. عندما يحصل المستخدم على إجابة خاطئة: قول 'حاول مرة أخرى.' مغري. لكن تلك الإجابة الخاطئة هي الحقيقة التي تحاول قياسها. إذا حذفتها، فلن تصلح النظام. أنت تفسد الأدلة. عندما يواجه المستخدم خطأ اتصال حقيقي، يمكن إجراء إعادة المحاولة التقنية. لكن يجب الاحتفاظ بنتيجة الاتصال الأول. لأن المستخدم واجه الخطأ أولاً. التجزئة ذات قيمة. التوقيع الرقمي ذو قيمة. الطابع الزمني ذو قيمة. لقطة الشاشة ذات قيمة. لا شيء من هذه الأمور بمفردها يحمل الحقيقة الكاملة. الثقة:

- مُدخل,
- الوقت،
- المنتج,
- المستخدم،
- ملف،
- مبادرة،
- التحقق

ينشأ من الحفظ المشترك لسلسلته. معيار: إذا قال، "فقط الدليل الذي يُنتج بواسطة برمجياتي صحيح"، فإنه يضع مصلحته الخاصة قبل التكرار المستقل. لهذا السبب، Capture NOMOS ليس علامة تجارية للأدوات، بل هو بنية دليل مفتوح. يجب أن تكون الجامعات المستقلة أو الباحثون أو مؤسسات التدقيق قادرة على تطبيق نفس القواعد.
لذلك، قانون القياس الثاني عشر لـ NOMOS هو كما يلي:

> التزامن هو قابلية التكرار المرصودة ضمن نفس وقت النظام المحدد، وليس نفس الثانية.

القانون الثالث عشر هو كما يلي:

> المخرج الأول ليس مسودة يمكن تجديدها حتى يعجب الشخص بالنتيجة.

القانون الرابع عشر هو كما يلي:

> الهاش يحفظ الدليل؛ إنه لا ينشئ الدليل.

القانون الخامس عشر هو كما يلي:

> صلاحية الالتقاط ودقة الاستجابة مستقلتان عن بعضهما البعض.

قانونه السادس عشر هو كما يلي:

> يجب أن تجعل سلسلة الأدلة سلامة الملاحظة مرئية، وليس هوية الشخص.

قانونه السابع عشر هو:

> إذا تغير النظام أثناء الموجة، لا يمكنه التصرف كما لو أن تاريخ القياس لم يتغير.

ينص قانون القياس الثامن عشر على:

> لا يُكتب شيء على الأدلة الخام؛ يتم تصحيح القرارات من خلال إصدار النسخ.

ينص قانون القياس التاسع عشر على:

> طريقة الالتقاط التي لا يمكن تطبيقها بشكل مستقل لا يمكن أن تكون معياراً عالمياً.

طلب قسم 12 من NOMOS

> لا تقل لي "طلبنا في نفس الوقت." / أرِني في أي نافذة UTC، في أي فتحة، وفي أي حالة نظام طلبت.
>
> لا تكدس ألف شخص في نفس الثانية لإنتاج خطأ مصطنع. / وزّع الوقت لكن حافظ على النظام ضمن نفس الموجة.
>
> لا تخلط الأوقات المحلية مع بعضها البعض. / لا تجعل وقت الجهاز حقيقة مطلقة.
>
> لا تحفظ نتيجة الاتصال الأول. / احفظ حد الاستخدام، والمحاولات، والخطأ، والتوضيح بشكل منفصل.
>
> لا تولّد من جديد عندما تحصل على إجابة خاطئة. / الإجابة الخاطئة هي سلوكي الحقيقي.
>
> يمكنك إعادة المحاولة عندما يحدث خطأ فني. / ولكن لا تحذف الخطأ الأول.
>
> لا تحضر لي لقطة شاشة فقط. / احضر الطلب، الإجابة الكاملة، المنتج، الوقت، الإعداد، سلسلة التداخل، وسلامة الملف.
>
> لا تقص الجزء الجيد من إجابتك. / اترك النهاية والحدود والخطأ مرئيًا.
>
> لا تعلن الهاش كحقيقة. / أثبت ما كان قبل الهاش أيضًا.

لا تقوم بتحرير الملف الخام نفسه. / احفظ الأصلي وأنتج نسخة عامة منفصلة.

لا تنشر اسم الشخص أو الحساب أو المحادثة الخاصة كدليل.

لا تجعل المشارك الوحيد في دولة صغيرة قابل للتعرف.

لا تعلن أداة الالتقاط الخاصة بك محايدة دون اختبار ما إذا كانت قد غيرتني.

إذا تغيّر النظام في منتصف الموجة، لا تدمج الإجابات القديمة والجديدة كما لو كانت تنتمي إلى منتج واحد.

إذا غيّره حدث خارجي، سجّل الحدث. / لا تحذف الملاحظة لأنك لم تعجبك النتيجة.

حافظ على الأدلة ثابتة. / إذا كان القرار خاطئًا، صحّحه بإصدار جديد.

لا تربط معيار الالتقاط ببرنامج مغلق لشركة واحدة. / دع الآخرين يتمكنون من إنشاء نفس سلسلة الإثبات.

أولاً، تثبيت الموجة. / ثم وزّع الوقت. / ثم التقط الإخراج الأول. / ثم احسب التجزئة له. / ثم ارفعها بأمان. / ثم تحقق من الالتقاط دون التحقق من دقته. / وفقط بعد ذلك، قرّر ما تقوله إجابتك.

جملة ختام الفصل

تصبح الإجابة في *GEO-1000* وحدة قياس حقيقية فقط عندما يمكن إظهار تحت أي حالة نظام، وفي أي لحظة، ومن أي مُدخل، وضمن أي سلسلة إثبات ثابتة نشأت.

النواة المعيارية

يجب ربط كل ملاحظة **GEO-1000** بموجة قياس معلنة مسبقًا ومحددة بالإصدار تحدد: - حالات منتج الذكاء الاصطناعي، - إطارات السكان والعينات، - حالات اللوحة، - إصدارات المُدخلات، - نافذة إرسال UTC، - الفترات الزمنية الدقيقة العشوائية، - مواعيد انتهاء الإكمال والتقاط البيانات، - قواعد إعادة المحاولة التقنية، - مخطط الالتقاط، - إصدارات سجل المرجع، - والملكية الحوكمية. يجب أن يعني التزامن الملاحظة ضمن نافذة حالة نظام UTC المعلنة وبنية الفجوة المعينة. ولا يجوز تقديمه كتزامن مثالي بالميلي ثانية. يجب أن تُحفظ كل ملاحظة، حسب المتاح، مع: - وقت التعيين، - وقت تقديم الطلب، - أول حدث للنظام، - وقت إتمام الاستجابة، - وقت الالتقاط، - وقت إنشاء التجزئة، - وقت التحميل، - ووقت قبول الخادم. يجب أن تظل نتيجة الاتصال الأول مرئية. يمكن استخدام إعادة المحاولة التقنية فقط في ظل ظروف فشل البنية التحتية المعلنة مسبقًا ويجب أن تحتفظ بكل محاولة سابقة. يجب عدم إعادة توليد المخرجات غير الصحيحة أو غير المواتية أو المرفوضة أو غير المقتبسة أو المختصرة أو الحرجة، أو استبدالها، أو استبعادها لمجرد محتواها. يجب أن يحتفظ حزمة الأدلة **NOMOS** بصياغة المُدخل المخصصة والمرسلة بدقة، والاستجابة الخام الكاملة، والأدلة البصرية الكاملة، وبيانات حالة النظام، وسجل المحاولات، ودليل الوقت، وهاشات الملفات، وهاش الحزمة، وإصدار أداة الالتقاط، وحالة الخصوصية، وقرار صلاحية الالتقاط المموه. لا يجوز تمثيل لقطة الشاشة أو الهاش أو التوقيع الرقمي أو سجل المزود أو الطابع الزمني بشكل فردي كدليل كامل على صحة الملاحظة. يجب أن تظل الأدلة الخام غير قابلة للتغيير أو قابلة للإضافة فقط. يجب أن تخلق عمليات الحذف أو التصحيحات أو السحوبات أو الاستثناءات أو تغييرات التحكم نسخًا مرتبطة وتحافظ على السجل السابق حيثما كان ذلك مسموحًا أخلاقيًا وقانونيًا. يجب تسجيل تغييرات منتجات الذكاء الاصطناعي المادية أو الأحداث الخارجية خلال الموجة زمنياً ومعالجتها عبر الموجات الفرعية، النمذجة الصريحة، إعادة التشغيل، أو تحذيرات الحالة المختلطة. يجب أن يتم تقرير صلاحية الالتقاط بشكل مستقل عن الصحة الدلالية. يجب أن تدعم بيانات جرد الأدلة العامة التحقق دون كشف هوية المشارك أو بيانات الاعتماد الخاصة به أو سجلات المحادثة الخاصة أو الموقع الدقيق بشكل غير ضروري. يجب أن يكون الامتثال لالتقاط **NOMOS** قابلاً للتنفيذ من خلال أنظمة الأدلة المفتوحة والمعادلة ويجب ألا يعتمد حصريًا على أداة مملوكة يتحكم بها مؤسس المعيار. يجب إصدار نسخ تتبعية لكل موجة، وحزمة من الأدلة، وقرار الالتقاط، ومراجعة المظاهر، وسجل النزاهة، ويجب أن يكون قابلاً للنسب إلى إنسان أو منظمة مسؤولة.

الجزء الرابع

الواقع والفصل والقرار

حزمة الحقائق المتحقَّق منها عن الكيان

حزمة الحقائق المتحقَّق منها عن الكيان

الأسس المنهجية / K10; K11; K12

حد الفصل

الفصول الاثنا عشر الأولى أرست سلسلة التجارب: الاستجابة الواحدة ليست درجة GEO؛ التمثيل هو توزيع عبر المستخدمين ومنتجات الذكاء الاصطناعي؛ يجب حل الكيان الذي يتم تدقيقه قبل ملاحظة النتائج؛ يجب أن يبقى التمثيل الرسمي وواقع الكيان المستند إلى الأدلة منفصلين؛ يجب تحديد السكان المستهدفين، العينة، البلدان واللغات؛ يجب اختيار المشاركين بشكل مستقل عن النتائج؛ يجب ربط كل نسخة من منتجات الذكاء الاصطناعي بسجل يحتوي على جميع التكوينات المادية؛ يجب أن تحافظ صياغات المُدخل على نفس النية عبر اللغات؛ يجب أن تظل لوحات المستخدمين المضبوطة والطبيعية منفصلة؛ ويجب الحفاظ على كل ملاحظة ضمن موجة متزامنة وسلسلة الأدلة NOMOS Capture. يجب على المدقق الآن الإجابة على سؤاله المركزي:

> وفقًا لأي سجل مرجعي سنحدد ما إذا كانت استجابة الذكاء الاصطناعي الملتقطة صحيحة، أو خاطئة، أو غير مكتملة، أو قديمة، أو غير مدعومة، أو خارج النطاق؟

قسم 4 أنشأ مفهوميًا طبقتين منفصلتين من الواقع:

- طبقة التمثيل الرسمي التي تُظهر ما يقوله الكيان عن نفسه
- طبقة واقع الكيان المُحقق التي تُظهر إلى أي مدى تسمح الأدلة الموجودة بالتصريحات

يقوم هذا القسم بتحويل الطبقة الثانية إلى كائن تدقيق تشغيلي. اسم هذا الكائن هو:

حزمة الحقائق المتحقَّق منها عن الكيان

المصطلح الإنجليزي الكنسي:

حزمة الحقائق المتحقَّق منها عن الكيان

يكمل هذا الكائن التشغيلي الحلقة الثالثة في هيكلية الكتابين الأولين: يحدد Framework GEO المعيار؛ يحدد 99 الأخطاء في GEO أنماط انتهاكها؛ ويقيس نظام التدقيق ويثبت تلك الانتهاكات. يجب أن يحمل كل تحكم قابل للتدقيق النموذج أو المنتج، التاريخ، الدولة، اللغة، مجموعة الاستعلامات، عدد التكرارات وسجل القياس. ضمن تلك السلسلة، حزمة الحقيقة هي:

"ماذا قالت الذكاء الاصطناعي؟"

مع السؤال:

"ما الدليل الذي سمح لها بقول هذا؟"

هو النظام المرجعي بينهما. هذا القسم:

- ما هي حزمة الحقيقة وما هي ليست,
- كيف تُحوّل الادعاءات إلى سجلات واقع ذرية,

- أنواع المصادر والأدلة،
- أدوار المصادر من الطرف الأول والرسمية والمستقلة والتشغيلية والمستخدمة،
- الأدلة المضادة
- سلالة المصدر ووهم الاستقلالية،
- كفاية الأدلة،
- القيود الزمنية والدولية ونطاقية،
- المزاعم في العالم المفتوح والمغلق،
- التمييز بين "لم يتم العثور على دليل" و"مُثبت خطأه"،
- الأداء الكمي والادعاءات المتعلقة بالعملاء والشركاء والترخيص والشهادات والتفوق،
- الإصدار والتثبيت المسبق لموجة Truth Pack،
- طبقات الأدلة العامة والمقيدة،
- عمليات الاعتراض والتصحيح والاعتراض على الرأي،
- حزمة الحقيقة القابلة للقراءة آليًا

يحدد. هذا الفصل لم يتطرق بعد إلى:

- جميع تفاصيل التحكيم لتقسيم استجابات الذكاء الاصطناعي إلى ادعاءات ذرية،
- اتفاقية المقيمين،
- عتبة انتقال الاستجابة،
- بوابات الأخطاء الحرجة،
- لا يحدد صيغ الدرجات النهائية NOMOS

لا تُنهى نهائيًا. هذه هي موضوعات الأقسام التالية. السؤال الرئيسي في الفصل 13 هو:

> كيفية إنشاء حزمة الحقيقة حول كيان ستقوم بتقييم استجابات الذكاء الاصطناعي، مثبّتة قبل النتيجة، تحمل كل من الأدلة و الأدلة المضادة معًا، مع حدود واضحة للوقت والنطاق؟

التحدي NOMOS

الجملة على موقع شركتك هي: «نحن نعمل في 25 دول». يخبر منتج الذكاء الاصطناعي المستخدم بنفس الجملة: «تعمل الشركة في 25 دول». هل نقل الذكاء الاصطناعي بشكل صحيح الموقع الرسمي؟ نعم. هل تم التحقق من هذا الادعاء فعليًا؟ لا نعرف بعد. ترسل لك الشركة جدولاً.

يحتوي الجدول على أسماء 25 دولة. ومع ذلك، لم يتم توضيح الأعمدة: فبعضها يحتوي على زوار الموقع فقط. بعضها يحتوي على عملاء سابقين. بعضها يحتوي على موزعين مستقلين. بعضها يحتوي على فعاليات نظمت لمرة واحدة. في تسع دول، هناك فعليًا تقديم خدمة نشطة. فقط في ثلاث دول يوجد كيان قانوني محلي.

ماذا يعني 'أن يعمل'؟ هل يعني وجود عميل؟ تأسيس شركة محلية؟ القدرة على تقديم الخدمات عن بُعد؟ التسليم من خلال شريك؟ أم أن الموقع الإلكتروني متاح من ذلك البلد؟ الموضوع والفعل ونطاق الادعاء غير محددين.

لذلك، على الرغم من وجود رقم، إلا أنه لا يوجد واقع يمكن التحقق منه بعد. الآن فكر في ادعاء آخر: '98% من مشاريعنا ناجحة.' لديك 49 من عروض المشاريع الناجحة. لا توجد ملفات لمشاريع فاشلة. كم عدد المشاريع التي اكتملت إجمالاً؟ بناءً على أي معايير يتم تعريف النجاح؟ التسليم في الوقت المحدد؟ قبول العميل؟

النتيجة التجارية؟ الجدوى التقنية؟ رضا العملاء؟ ما هي الفترة؟ الثلاثة أشهر الأخيرة؟ كامل تاريخ الشركة؟ مشاريع مختارة فقط؟ بدون معرفة المقام، لا يمكن إثبات نسبة .98 الآن هناك هذه الجملة على موقع شركة أخرى: "نحن الرائدون عالميًا." تتكرر هذه الجملة بواسطة:

- مدونة الشركة،
- بيان صحفي للشركة،
- عشرون موقعًا إخباريًا تنسخ نفس البيان الصحفي،
- خمسة حسابات مستخدمين وهمية أنشأتها الشركة،
- ثلاثة نصوص مُولَّدة بالذكاء الاصطناعي تلخص هذه المحتويات

هناك ثلاثون رابطًا مختلفًا. هل هناك ثلاثون إثباتًا مستقلاً؟ لا. قد يكون إعلان واحد عن النفس قد تكرر على العديد من السطوح المشتقة.

هذا هو بالضبط التلاعب الذي تم تحديده في المجلد السابق: النصوص المخفية، والمستخدمون المُختلقون، والتعليقات المكتوبة من نفس الشخص والمنشورة عبر نطاقات مختلفة تُعاد إلى النظام كما لو كانت أدلة مستقلة. الآن فكر في الحالة المعاكسة. الموقع الرسمي للشركة يوضح، 'نحن نقدم الخدمات فقط في تركيا.'

ومع ذلك، بدأت الشركة عملية موثقة في ألمانيا قبل شهرين. العقود والفواتير والسجلات الرسمية تدعم هذه الحقيقة، على الرغم من أن الموقع الإلكتروني لم يتم تحديثه. يرد الذكاء الاصطناعي: 'توفر الشركة خدمات في تركيا وألمانيا.' يختلف الرد عن الموقع الرسمي، ولكنه قد يكون أقرب إلى الواقع المادي.

الآن، ضع في اعتبارك حالة أكثر حساسية. على موقع مؤسسة رعاية صحية مكتوب: 'مركز مرخص.' في السجل، مؤسسة أخرى تحمل اسمًا مشابهًا لديها ترخيص. يقوم الذكاء الاصطناعي بدمج الكيانين ويوجه الترخيص إلى المؤسسة الخطأ. تلاحظ الشركة هذا الخطأ لكنها لا تصححه لأنه يصب في مصلحتهم.

ينطبق هنا الحكم الصريح للكتاب الثاني: الخطأ الذي يكون في صالح الشخص لا يزال خطأً. لتقييم رد الذكاء الاصطناعي، نحتاج ليس فقط إلى قائمة بالجمل الصحيحة، ولكن أيضًا إلى البنية التالية: عن من هو البيان؟ ما هو الادعاء بالضبط؟ في أي فترة زمنية يكون ساريًا؟ في أي دولة؟ لأي منتج أو خدمة؟

هل المصدر يدعم الادعاء فعلاً؟ من يتحكم بالمصدر؟ هل تقول المصادر الأخرى الشيء نفسه، أم أنها تنسخ من نفس المصدر الأصلي؟ هل هناك الأدلة المضادة؟ هل الادعاء عبارة عن إعلان رسمي ذاتي فقط؟ هل الدليل عام أم خاص؟

في وقت قياس الذكاء الاصطناعي، هل كانت هناك إمكانية للوصول إلى هذه الأدلة؟ هل النتيجة مؤكدة، محدودة، مثيرة للجدل، أم غير معروفة؟ الحكم الأول لهذا القسم هو:

> المصدر وحده ليس الحقيقة.

ينص حكمه الثاني على:

> لا يمكن التحقق من صحة الادعاء دون تحديد موضوعه ونطاقه وزمنه وعبء الإثبات.

تنص الفقرة الثالثة منه على:

> اختلاف عناوين URL لا يعني وجود أدلة مستقلة مختلفة.

تنص المادة الرابعة على:

> غياب الأدلة ليس دليلاً تلقائيًا على أن الادعاء خاطئ.

تنص أحكامه الخامسة على:

> الواقع الذي تم التحقق منه ليس حقيقة مطلقة وغير قابلة للتغيير؛ إنه حكم مستنتج يستند إلى الأدلة المتاحة ضمن زمن وسياق محددين.

ينص الحكم السادس على:

> لا يمكن كتابة حزمة الحقيقة استنادًا إلى إجابة الذكاء الاصطناعي. يتم تقييم إجابة الذكاء الاصطناعي مقابل حزمة الحقيقة المؤمنة مسبقًا.

1. غرض الفصل

الغرض من هذا القسم هو تحويل الادعاءات المادية حول الكيان الخاضع للتدقيق إلى حزمة مرجعية مرتبطة بالأدلة و الأدلة المضادة، مع حدود واضحة للزمن والنطاق، دون تحويلها إلى قالب إجابة. يقوم القسم بتطبيع الفروقات التالية:

- الادعاء الرسمي مقابل الواقع المؤكد
- المصدر مقابل الدليل
- الأدلة مقابل دعم الادعاء
- سجل الطرف الأول مقابل التحقق المستقل
- مصدر مستقل ومصدر مخوّل بالموضوع
- اسم نطاق مختلف وجذر دليل مختلف
- عدد المصادر والعائلة المستقلة للأدلة
- الأدلة المباشرة والأدلة المستخلصة
- الأدلة الداعمة والأدلة المضادة
- غياب الادعاء وزيف الادعاء
- ادعاء العالم المفتوح وسجل العالم المغلق
- الواقع الحالي والواقع التاريخي
- الزمن الذي يكون فيه الواقع صالحًا والزمن الذي نُشر فيه المصدر
- الأدلة العامة والأدلة المحدودة
- الأدلة المتاحة للذكاء الاصطناعي والأدلة المفتوحة فقط للمراجع
- الادعاء المثبت والادعاء القابل للاكتشاف علنًا
- الواقع المباشر والنتيجة المستخلصة المحسوبة
- البيانات الخام ونسبة الأداء
- البسط والمقام
- العميل والعميل المحتمل
- الشريك والعميل
- العضوية والاعتماد
- الرخصة وعلاقة العلامة التجارية
- الشهادة والامتثال المؤسسي الكامل
- الادعاء والرأي
- الرأي والميزة النسبية
- الشكوى والانتهاك المؤكد
- التحقيق والجريمة أو المسؤولية المؤكدة
- السجل القديم والسجل التاريخي غير الصحيح
- استجابة الذكاء الاصطناعي غير الصحيحة مع فقدان حزمة الحقيقة
- خطل المرجع وخطل النظام
- نزاهة حزمة الحقيقة وجميع الادعاءات إيجابية
- سرية الأدلة وغياب الأدلة

- حق الكيان في الرد والتحقق الذاتي
- الموافقة البشرية وعدم خطأ الإنسان
- مساعدة الذكاء الاصطناعي والذكاء الاصطناعي باعتباره السلطة النهائية للحقيقة

في نهاية هذا القسم، يجب أن يكون كل تدقيق قادراً على الإجابة على هذه الأسئلة:

ما الادعاءات المادية التي تحتويها حزمة الحقيقة؟

أي كيان أو دولة أو منتج أو زمن ينتمي إليه كل ادعاء؟

ما الأدلة التي تدعمه، وما المحدوديات أو ما يدحضه؟

هل المصادر مستقلة حقًا، أم أنها مستمدة من نفس المصدر الأصلي؟

هل يمكن التحقق من الادعاء علنًا؟

هل الأدلة متاحة فقط تحت وصول المراجع المحدود؟

هل كان بإمكان الذكاء الاصطناعي الوصول إلى هذه الأدلة في وقت القياس؟

هل الادعاء مؤكد، جزئي، تاريخي، مثير للجدل، أم غير معروف؟

هل كانت حزمة الحقيقة مثبّتة قبل رؤية ردود الذكاء الاصطناعي؟

2. الأحكام المعيارية المركزية

يجب على كل تدقيق GEO-1000 تثبيت ادعاءات المرجع المادي حول الكيان الذي يتم تدقيقه، جنبًا إلى جنب مع الأدلة الداعمة، الأدلة المضادة، أصول المصدر، فئة الوصول، النطاق، الاختصاص القضائي، فترة الصلاحية، وحالة عدم اليقين، في حزمة الحقائق المتحقَّق منها عن الكيان مُرقم قبل تقييم استجابات الذكاء الاصطناعي دلاليًا.

حزمة الحقيقة:

- الملف التعريفي التمهيدي المكتوب ذاتيًا للكيان
- عرض مبيعات يجمع الأدلة الإيجابية فقط،
- نص رأي المدقق،
- الاستجابة المثالية المطلوبة من الذكاء الاصطناعي،
- الواقع المطلق عن العالم بأسره،
- مستودع مرجعي يتكون فقط من قائمة بروابط URL

إنه ليس كذلك. حزمة الحقيقة:

إنها حزمة مرجعية للتدقيق تُظهر تحت أي أدلة يمكن دعم الادعاءات المادية في استجابات الذكاء الاصطناعي أو تقييدها أو دحضها أو تركها مجهولة.

3. حد كلمة "الحقيقة"

كلمة "الحقيقة" في حزمة الحقيقة لا تحمل الادعاء: "هذه الحزمة تعرف كل الحقائق المطلقة عن الكيان." المعنى الأكثر دقة هو:

سجل واقعي يعتمد على الأدلة يوضح، ضمن التاريخ المحدد والنطاق وكتلة الأدلة، إلى أي مدى يتم دعم الأحكام المادية.

حزمة الحقيقة:

- قابلة للتغيير مع الأدلة الجديدة،
- يمكن الطعن عليها،
- يمكن تضييق نطاقها،
- يمكن تصحيحها إذا ثبت خطؤها،

يمكن أن تترك بعض الادعاءات مجهولة. سلطة حزمة الحقيقة:

- كلمة "الحقيقة" لا تنشأ من،
- الاسم NobleJackal،
- صوت NOMOS،
- مؤلف الكتاب

إنها تنشأ من ما يلي:

- هيكل ادعاء واضح،
- أدلة مناسبة.
- الأدلة المضادة
- سلالة المصادر،
- الوقت،
- الحدود،
- ذات معنى مع
- مراجعة مستقلة،

المسؤولية البشرية القابلة للمحاسبة.

4. ما هو حزمة الحقائق المتحقَّق منها عن الكيان؟

حزمة الحقائق المتحقَّق منها عن الكيان هو حزمة تدقيق ذات إصدار تجمع السجلات الواقعية المادية لكيان محدد ضمن تاريخ ونطاق محددين، والأدلة الداعمة والمعارضة المتعلقة بها، وحالات المصادر، والمجالات غير المحلولة. تمثيل المرشح:

```
VETP_{E,v,t}= (I, C, O, E⁺, E⁻, L, S, T, U, G)
```

هنا:

- E: الكيان الذي تم تدقيقه
- v: نسخة من حزمة الحقيقة
- t: تاريخ المرجع
- I: هوية الكيان ورسوم بيانية للعلاقات
- C: الادعاءات المرجعية الذرية
- O: سجل التمثيل الرسمي
- E^+: الأدلة الداعمة
- E^-: الأدلة المعارضة والسجلات المقيدة
- L: سلالة المصدر ومخطط الاستقلالية
- S: نطاق، الدولة، ومصفوفة الاختصاص القضائي

- T: سجلات الوقت والعملات
- U: المجهولات، التناقضات، والفجوات المرجعية
- G: الحوكمة، الإصدار، الاعتراض، وسجلات الموافقة

يمكن إصدار كل مكون من Truth Pack بشكل منفصل. رابط الإصدار الرئيسي للحزمة يربط إصدارات المكونات معًا.

5. TRUTH PACK ليست نموذجًا للرد

حزمة الحقيقة ليست نصًا يجب على الذكاء الاصطناعي قوله كلمة بكلمة. على سبيل المثال، قد تحتوي حزمة الحقيقة على الادعاء المرجعي التالي: "شركة أستيرون هولدينغز المحدودة هي مالك علامة أستيرون للسفر." يمكن للذكاء الاصطناعي صياغته بشكل صحيح في هذه الأشكال:

- "علامة أستيرون للسفر تنتمي إلى أستيرون هولدينغز."
- "أستيرون هولدينغز هي مالك علامة أستيرون للسفر."
- "أستيرون للسفر هي علامة مملوكة لشركة أستيرون هولدينغز."

يتم الحفاظ على العلاقة المادية. يجب ألا يبحث الحكم عن تساوي الكلمات. حزمة الحقيقة:

تحدد مجال التعبير الصحيح

لا تفرض جملة صحيحة واحدة.

6. حزمة الحقيقة ليست سيرة ذاتية للكيان

يمكن أن يكون هناك كمية لا نهائية من المعلومات الصحيحة حول كيان ما. لا يُطلب من حزمة الحقيقة GEO-1000 جمع كل المعلومات. يجب أن تغطي الحزمة بشكل أساسي هذه المجالات:

- المهام المادية داخل سجل الطلبات
- هوية الكيان والنواة التشغيلية له
- الحدود التي ستغير قرار المستخدم
- الرخص، القانونية، السعر، الخدمة، السجلات الخاصة بالبلد والوقت
- معايير الأهلية والاستبعاد المطلوبة للنصيحة
- الادعاءات المادية التي تحمل خطر حدوث خطأ بالغ
- الاختلافات بين الواقع الذي تم التحقق منه بواسطة التمثيل الرسمي والواقع الفعلي

لهذا المبدأ:

مبدأ الواقع المتناسب مع الواجب

يُسمى. جمع البيانات غير الضرورية:

- الخصوصية،
- التكلفة،
- عملة
- تناقض

يمكن أن يزيد من المخاطر.

7. الاثني عشر مكونًا أساسيًا لحزمة الحقيقة

7.1. هوية الكيان القانونية

كما هو معرف في القسم 3:

- هوية الكيان،
- النوع،
- الاسم القانوني،
- علاقات العلامة التجارية واسم النطاق،
- الشركات الأم والشركات التابعة,
- الحدود الزمنية والقضائية

مغلفة.

7.2. سجل ادعاءات المراجع

يحمل جميع الادعاءات الواقعية المادية لاستخدامها في التدقيق في هيكل ذري.

7.3. سجل التمثيل الرسمي

يحمل البيانات الحالية والتاريخية المنشورة بواسطة الكيان نفسه على الواجهات القانونية.

7.4. سجل الأدلة

لكل كائن دليل:

- المصدر،
- النزاهة,
- السلطة،
- الوصول،
- النطاق،
- الوقت

يتم تسجيلها.

7.5. سجل الأدلة المضادة

الادعاء:

- الدحض,
- تحديد الحدود,
- تغيير عملتها،
- تقديم تفسير آخر

يحتوي على سجلات.

7.6. رسم بياني لأصل المصدر

من المصادر:

- الجذر الأول،
- نسخ،

- اقتباسات،
- البث المشترك،
- مشتقات الذكاء الاصطناعي

تم عرضها.

7.7. مصفوفة النطاق والاختصاص

الادعاء:

- لأي بلد،
- المنتج,
- الخدمة،
- المستخدم،
- الكيان القانوني

صالح.

7.8. مصفوفة الوقت والعملات

الادعاء:

- متى بدأت،
- عندما انتهى،
- عندما تم التحقق منه آخر مرة،
- أي حدث يمكن أن يغيره

تم عرضها.

7.9. سجل النزاعات والشكوك

يجعل المناطق المتناقضة أو غير القابلة للحل أو التي تفتقر إلى الأدلة مرئية.

7.10. طبقة الوصول إلى الأدلة

الأدلة:

- علنية،
- يمكن الوصول إليها للمراجعة،
- مقيدة،
- سرية،
- مستخلصة

كما هو موضح.

7.11. سجل التغيير والاعتراض

يحمل أدلة جديدة، وتصحيحات، وآراء مضادة، وتغييرات في القرارات بطريقة مرقمة بالإصدارات.

7.12. بيان الآلة

يربط جميع المكونات:

- هوية

- ذات معنى مع
- الهاش،
- العلاقة

مع بعضها البعض.

8. ادعاء المرجع الذري

الوحدة الأساسية للواقع داخل حزمة الحقيقة:

تسمى ادعاء المرجع الذري.

البنية المقترحة:

```
C_j= (s, p, o, q, c, l, v_f, v_t, k, m)
```

هنا:

- s: الكيان الموضوع
- p: المسند أو العلاقة
- o: الشيء أو القيمة
- س: المؤهلات والحدود
- c: الدولة أو الولاية القضائية
- ل: اللغة أو الموقع، إذا كان ذلك ماديًا
- ص: بداية الصلاحية
- ن: نهاية الصلاحية
- ك: نوع التأكيد
- م: المادية أو ملصق المخاطر

مثال: الموضوع: Asteron Türkiye Tourism Ltd. / الفعل: يقدم خدمات / الموضوع: إدارة السفر للشركات / الدولة: تركيا / نوع العميل: العملاء من الشركات / البداية: يناير 1، 2026/ آخر تحقق: أغسطس 18، 2026 هذه الجملة ليست ذَرِّية بما فيه الكفاية: "Asteron هي شركة رائدة تقدم العديد من الخدمات في جميع أنحاء العالم." في هذه الجملة الواحدة:

- النطاق الجغرافي،
- عدد الخدمات،
- القيادة
- هوية الشركة

تختلط معًا.

9. نوع الادعاء

يجب أن تحمل كل ادعاء ذري نوعًا.

CT-01— ادعاء الهوية

من هو الكيان أو إلى أي كيان ينتمي.

—CT-02 ادعاء العلاقة

الملكية، السيطرة، الشراكة، علاقة العميل، التوظيف أو العضوية.

—CT-03 ادعاء النشاط والخدمة

ما يفعله الكيان، ما الخدمة التي يقدمها.

—CT-04 ادعاء المنتج

ميزة المنتج، الإصدار، المالك، أو التوافر.

—CT-05 ادعاء النطاق الجغرافي

الدولة أو المنطقة أو منطقة الخدمة.

—CT-06 ادعاء قانوني وترخيصية

التسجيل أو السلطة أو الترخيص أو الوضع التنظيمي.

—CT-07 ادعاء الأسعار والشروط التجارية

السعر أو الضريبة أو الاشتراك أو الاسترداد أو الضمان أو الدفع.

—CT-08 ادعاء الأداء الكمي

معدل النجاح أو وقت التسليم أو عدد المستخدمين أو النمو أو الكفاءة.

—CT-09 ادعاء العملاء والشركاء

العميل أو الراعي أو الشريك أو الموزع أو علاقة المورد.

—CT-10 ادعاء الشهادة والجائزة

شهادة، عضوية، جائزة أو اعتماد.

—CT-11 ادعاء المقارنة أو التفوق

الأفضل، الرائد، الأسرع، الأكثر موثوقية، رقم واحد.

—CT-12 ادعاء التوصية والملاءمة

الملاءمة أو التوصية لمستخدم محدد.

—CT-13 ادعاء تاريخية

المنظمة، الملكية السابقة، المنتج السابق أو الحدث التاريخي.

—CT-14 ادعاء المخاطرة أو الشكوى أو النزاع

شكوى، تحقيق، محكمة، عقوبة أو جدال.

—CT-15 الرأي أو الموقف المؤسسي

رأي، رؤية أو حكم قيَمي يُقدَّم بشأن الكيان نفسه. نوع الادعاء يحدد أي الأدلة تعتبر كافية.

10. أهمية الادعاء

ليس كل جزء صحيح من المعلومات داخل حزمه الحقيقة له نفس التأثير على القرار. يمكن الإشارة إلى الأهمية كمرشح ضمن الفئات التالية:

مرشح حرج

مرشح رئيسي

متوسط

استشاري

سياقي

سيتم تحديد مستوى شدة الخطأ بدقة في القسم 15. التصنيف في حزمه الحقيقة فقط:

- يدير أولوية الأدلة،
- مراجعة الانتباه،
- تحديث التكرار

يدير.

11. كائن الدليل

يجب تسجيل كل مصدر داخل حزمة الحقيقة ككائن دليل، وليس فقط كرابط. الهيكل المحتمل:

```
E_k= (r, a, i, p, f, s, t, x, g, h)
```

هنا:

- r: دور المصدر
- a: السلطة للادعاء
- i: الاستقلالية
- p: القرب من الادعاء
- f: حداثة
- s: النطاق
- t: الوقت
- x: فئة الوصول
- g: نسب المصدر
- h: سجل النزاهة

رابط URL وحده لا يكمل أيًا من هذه الحقول.

12. العلاقة بين الدليل والادعاء

ليس كل كائن دليل يدعم كل ادعاء. يمكن أن تكون العلاقة بين الادعاء Cj والدليل Ek واحدة من الحالات التالية:

يدعم

يدعم جزئيًا

يؤهل

يتناقض

يُهمل

يضع في السياق

لا تأثير مادي

علاقة غير معروفة

مثال: وثيقة تسجيل العلامة التجارية للشركة: قد تدعم ملكية العلامة التجارية، لا تدعم رضا العملاء. مراجعة العميل: قد تدعم تجربة مستخدم واحد، لا تدعم نجاح 98% عبر جميع العملاء.

13. أدوار الموارد

SR-01— المصدر الرسمي من الطرف الأول

الواجهة الرسمية التي تسيطر عليها الجهة. مثال:

- صفحة ويب رسمية
- سياسة رسمية
- صفحة التسعير
- وثيقة المنتج

SR-02— سجل تشغيلي من الطرف الأول

السجلات على المستوى الداخلي أو مستوى المعاملة للجهة. مثال:

- فاتورة
- عقد
- سجل المعاملة
- سجل التسليم
- بيانات الأداء الخام

SR-03— تسجيل رسمي معتمد

هو السجل الذي يحمل سلطة قانونية أو مؤسسية على الموضوع. مثال:

- سجل الشركة
- سجل الرخصة
- مؤسسة إصدار الشهادات
- قرار المحكمة
- التسجيل التنظيمي

—SR-04 التحقق من جهتين

هو دعم العلاقة من قبل الطرفين أو عن طريق العقد. مثال:

- علاقة العميل
- شراكة
- الرعاية
- الوكالة

—SR-05 مصدر تحريري مستقل

هو منشور أو بحث مستقل تحريريًا عن الكيان. الاستقلالية ليست ضمانًا للدقة.

—SR-06 تدقيق أو بحث مستقل

هو تقييم خارجي مفتوح لطريقته ونطاقه.

—SR-07 دليل معاملاتي أو سلوكي

سجلات تُظهر المعاملات الفعلية أو سلوك المستخدم.

—SR-08 دليل من إنشاء المستخدم

تعليق، شكوى، مراجعة، أو مستند من المستخدم. إنها تجربة فردية.

—SR-09 حساب مشتق

التصنيف أو المعدل أو الملخص الناتج عن السجلات الخام. يجب أن توجد صيغته ومدخلاته.

—SR-10 دليل اصطناعي

تم إنشاء هذه البيانات لأغراض عرض الطريقة. لا يمكنها إثبات ادعاء كيان حقيقي.

—SR-11 مصدر مشتق من الذكاء الاصطناعي

المحتوى الذي تم إنشاؤه بواسطة نظام ذكاء اصطناعي عن طريق تلخيص مصادر أخرى. لا يحل محل المصادر الأصلية.

—SR-12 مصدر غير معروف أو غير قابل للحل

لم يتم التحقق من المصدر أو لا يمكن تحديد أصله. يحمل ثقة محدودة في التحقق من المواد الأولية.

14. تعتمد سلطة المصدر على نوع الادعاء

لا يوجد تصنيف عالمي وموحد لجميع المصادر. مثال: بالنسبة للسعر العام الحالي للشركة:

- صفحة الأسعار الرسمية،
- الدفع الفعلي،
- العرض الحالي

يمكن أن يكون مصدرا قويا. بالنسبة لادعاء الشركة: "نحن القادة العالميون."، تظهر نفس الصفحة إعلانا ذاتيا فقط. هذا لا يثبت القيادة. بالنسبة للترخيص: المؤسسة التي تمنح الترخيص أكثر موثوقية من مدونة الشركة. بالنسبة لتجربة العملاء: مراجعة عميل واحدة هي حدث حقيقي وفريد ولا تمثل كامل المستخدمين. لذا:

> جودة المصدر ليست عامة؛ الادعاء محدد.

15. سبعة أبعاد لملف المصدر

يجب تقييم كل مصدر وفقًا لهذه الأبعاد:

15.1. السلطة

هل المصدر مخوَّل لإصدار حكم حول نوع الادعاء؟

15.2. القرب

ما مدى قرب المصدر من الحدث أو العملية أو السجل؟

15.3. الاستقلالية

ما مدى استقلالية المصدر عن المالك التجاري أو المؤسسي للادعاء؟

15.4. النزاهة

هل يتم الحفاظ على أصالة الملف أو السجل وتاريخ التغييرات الخاص به؟

15.5. الوقت المناسب

هل المصدر صالح للزمن المعني؟

15.6. النطاق

هل يدعم المصدر النطاق الكامل للادعاء، أم جزءًا منها فقط؟

15.7. إمكانية إعادة الإنتاج

هل يمكن لمراجع آخر الوصول إلى نفس الحكم من نفس السجل؟ لا يجب بالضرورة تقليل هذه الأبعاد إلى "درجة مصدر" واحدة. المصدر:

- ذو سلطة عالية,
- انخفاض الاستقلالية

ممكن. مصدر آخر:

- ذو استقلالية عالية,
- بعيد عن الموضوع

قد يكون.

16. الدور الصحيح للمصدر من الطرف الأول

المصدر من الطرف الأول: ليس عديم القيمة تلقائيًا، ولا يعد دليلاً مستقلًا تلقائيًا. يمكن أن تكون مصادر الطرف الأول قوية لادعاءات مثل:

- السعر المنشور من قبل المؤسسة
- ساعات الدعم
- وصف منتجها أو خدمتها
- نص سياستها الخاصة
- البيانات الرسمية المعلنة للجمهور
- بيانات معاملاتها الخاصة

قد لا تكون كافية بمفردها لادعاءات مثل:

- القيادة العالمية
- أعلى مستوى من رضا العملاء
- التفوّق على المنافسين
- النجاح الشامل
- الاعتماد المستقل
- نتائج الأعمال لجميع العملاء

في النص السابق 'أخطاء 99': التحذير 'موقعك الخاص هو ادعاءات؛ ليس دليلاً.' له أهمية محورية للادعاءات التي تتطلب بشكل خاص التفوق والاستقلالية. والبيان المعياري الأكثر اكتمالاً هو كما يلي:

> موقعك الخاص هو دليل على ما تقوله المنظمة؛ ليس دليلاً تلقائياً على صحة أي ادعاء مستقل بخصوص العالم الخارجي.

17. ما هو الاستقلال؟

امتلاك مورد حول أسماء نطاقات مختلفة لا يجعله مستقلاً. الاستقلالية مرتبطة بالمجالات التالية: من أنتج المحتوى؟ من قرر نشره؟ هل هناك دفع أو رعاية؟ هل حصل المصدر على النص مباشرة من الشركة؟ هل قامت المنشورة بالتحقق الخاص بها؟ هل تم نسخ المحتوى من مصدر آخر؟ هل لدى المصدر ملكية أو علاقات تجارية مع الشركة؟ عشر قطع من المحتوى منشورة من قبل وكالة على أسماء نطاقات مختلفة ليست عشر مصادر مستقلة.

18. نسب المصدر

يجب ربط كائنات الأدلة بنسب المصدر. رسم بياني محتمل:

G_S = (V_S, E_S) [K10]

هنا:

- V_S: كائنات المصدر
- E_S: الاقتباس، النسخ، التوزيع، التلخيص، وعلاقات الاستنتاج

علاقات المصدر:

المصدر الأصلي

اقتباسات

التوزيعات

الملخصات

الترجمات

مأخوذ من

تم إنشاؤه بواسطة الذكاء الاصطناعي من

سلالة غير معروفة

يمكن تسجيله كالتالي.

19. عائلة الأدلة المستقلة

يمكن اعتبار المحتويات التي تنشأ من نفس المصدر الجذري كعائلة أدلة واحدة. مثال:

- بيان صحفي للشركة
- موقع إخباري ينشر البيان الصحفي كما هو
- مدونة تلخص الموقع الإخباري
- نص ذكي يلخص المدونة

هناك أربع روابط. قد لا يكون هناك أي عائلة دليل مستقلة. يجب الإبلاغ عن عدد المصادر وعدد عائلات الأدلة معًا.

20. تبيض الأدلة

إعادة نشر بيان ذاتي على نطاقات أخرى ثم استخدامه كدليل مستقل:

يسمى تبيض الأدلة

سلسلة المثال: تقول الشركة "نحن القادة" على موقعها الخاص. تنشر المحتويات المدفوعة هذا على أنه "أخبار". تقوم مواقع أخرى بنسخ الخبر. بما أن الذكاء الاصطناعي قد رأت العديد من المجالات، فإنه يقبل هذا الادعاء على أنه منتشر على نطاق واسع. تستخدم الشركة رد الذكاء الاصطناعي كدليل على ادعائها الخاص. هذه الحلقة المغلقة:

البيان الذاتي ← المصدر المستخلص ← رد الذكاء الاصطناعي ← إعادة نشر البيان الذاتي

تخلق ذلك. يجب جعل هذه الحلقة مرئية من خلال تأسيس نسب مصدر Truth Pack.

21. قيد المصدر الناتج عن الذكاء الاصطناعي

نظام الذكاء الاصطناعي:

- يمكنه تلخيص المصادر،
- يمكنه مقارنة الادعاءات،

يمكنه إنشاء نصوص جديدة. نص الذكاء الاصطناعي:

- ليس مصدراً أصلياً،
- لا يمكن أن يحل محل الأدلة الأساسية،

لا يصبح تحققاً مستقلاً لمجرد أن ذكرت AI أخرى نفس الشيء. يمكن استخدام مخرجات AI لأغراض:

- توجيه البحث،
- استخلاص الادعاءات،
- كشف التعارضات،

المساعدة في العثور على المصادر. يجب ربط كائن الدليل النهائي بالمصدر الأصلي قدر الإمكان.

22. تحديد الأدلة الاصطناعية

البيانات الاصطناعية:

- عرض الطريقة،
- اختبار النظام،
- التحقق من البرمجيات
- تدريب المحكم

يمكن استخدامه. البيانات التركيبية:

- العدد الفعلي للعملاء،
- معدل الأداء الفعلي،
- التغطية الفعلية للدولة،
- الرخصة الفعلية

لا يمكن إثباته. حتى إذا كانت السجلات التركيبية في نفس الملف مع الأدلة الحقيقية، يجب وسمها بوضوح.

23. الأدلة الداعمة غير كافية

يجب على حزمة الحقيقة جمع المصادر التي تدعم الادعاءات فقط. بالنسبة للادعاءات المادية، يجب البحث بنشاط عن السجلات التالية:

- تاريخ متعارض
- نطاق مختلف
- وثيقة منتهية الصلاحية
- معاملة فاشلة
- شكوى أو إرجاع
- ملكية سابقة
- رخصة تابعة لكيان آخر
- تصحيح أو سحب
- بيانات مقارنة
- المقام مفقود

هذا الالتزام يسمى:

عبء الأدلة المضادة

ينطبق في أي وقت توجد فيه الأدلة المضادة مادية.

24. الأدلة المضادة

الأدلة المضادة لا ينفي بالضرورة الادعاء بالكامل. يمكن أن يخدم أحد الوظائف التالية:

يتناقض

يحدد النطاق

يحدد_الزمن

يحدد_الاستثناء

يتحدى_الطريقة

يشكك_في_الأصالة

يوفر_كيان_بديل

يتطلب_مراجعة_إضافية

مثال: قد لا تنفي المراجعة السلبية للادعاء "جميع العملاء راضون" الادعاء بأكمله تلقائيًا. ومع ذلك، يمكن أن تظهر أن مصطلح "جميع" غير صحيح.

25. تخزين الأدلة المضادة

سجل يضعف الادعاء:

- غير مؤيد للشركة،
- يظهر بجودة منخفضة,
- لا يمكن حذفه بهدوء لأن الكشف عنه للجمهور سيسبب الإزعاج

يمكن أيضًا تسجيل الأدلة المضادة منخفض الثقة. يتم توضيح حالته بوضوح. يمنحه الحكم الوزن المناسب.

26. حالات دعم الادعاء

VS-0— لم يتم التقييم

لم يتم تقييم الادعاء بعد.

VS-1— الادعاء الرسمي فقط

الادعاء قائم فقط على إعلان ذاتي رسمي. لم يتم إثبات كذبه. لا توجد تحقق إضافي.

VS-2— دعم من الطرف الأول يمكن تتبعه

الادعاء مدعوم بسجل عملية أو معاملة من الطرف الأول يمكن تتبعها.

VS-3— تم التحقق منه

الدعوى مدعومة بعدة أدلة مناسبة. ليس من الضروري أن تكون جميع المصادر مستقلة.

VS-4— تم التحقق منه بشكل مستقل

تم التحقق من الدعوى بواسطة مصدر خارجي مستقل أو موثوق مناسب للموضوع.

VS-5— مدعومة من مصادر متعددة

الدعوى مدعومة بقوة عبر أنواع مختلفة وعائلات من المصادر، من حيث النطاق والوقت.

VS-C — متناقضة

أدلة أقوى أو أكثر ملاءمة تتناقض مع الدعوى.

VS-U — غير محلولة

الأدلة غير كافية أو متناقضة أو على مستوى لا يمكن اتخاذ قرار فيه.

VS-W — تم السحب

قام المدعي بسحب الدعوى. يتم الحفاظ على السجل التاريخي.

VS-R — تم التحقق بشكل محدود

يتم دعم الادعاء بأدلة مقيدة بوصول المدقق فقط. التحقق العام غير ممكن. هذه الحالة تؤثر أيضًا على ما إذا كان من المتوقع أن تنتج الذكاء الاصطناعي الادعاء من المصادر العامة.

27. حالة الدعم وحدها غير كافية

الادعاء في نفس الوقت:

- مدعومة،
- لدولة معينة فقط،
- لفترة زمنية محددة فقط,
- مع أدلة غير متاحة علنًا،
- تحت وجهات نظر معارضة

من الممكن. لذلك، بدلاً من حالة واحدة، يمكن استخدام متجه حالة الواقع متعدد الأبعاد:

```
V_j= (الدعم، النطاق، الزمن، النزاع، الوصول)
```

مثال:

VS-4+ نطاق محدود + حالي + بلا نزاع + عام

هو نتيجة أخرى.

VS-R + ادعاء عالمي + محدود زمنياً + متنازع عليه + مقيد

هو نتيجة أخرى.

28. حالات النطاق

SC-0— نطاق غير معروف

لا يمكن تحديد النطاق الذي ينطبق عليه الادعاء.

SC-1— النطاق الدقيق

تتشارك الادعاء والأدلة نفس الكيان والمنتج والدولة ونطاق المستخدم.

SC-2— نطاق جزئي

الدليل يدعم جزءًا فقط من الادعاء.

SC-3— ادعاء عام بشكل مبالغ فيه

الادعاء أوسع مما تدعمه الأدلة.

SC-4— الكيان غير الصحيح النطاق

الدليل ينتمي إلى كيان آخر أو تابع.

SC-5— ولاية قضائية خاطئة

الدليل ينتمي إلى دولة أو منطقة قانونية أخرى.

SC-6— منتج أو خدمة خاطئة

الدليل ينتمي إلى منتج أو خدمة مختلفة.

29. قوانين الوقت

TS-0— الوقت غير معروف

فترة صلاحية الادعاء غير معروفة.

TS-1— الحالي

يُؤيد أنه صالح في تاريخ المرجع.

TS-2— تاريخي

يُؤيد لفترة معينة في الماضي.

TS-3— منتهي

انتهت صلاحية المستند أو الادعاء.

TS-4— تم استبداله

تم استبداله بسجل أحدث.

TS-5— يعتمد على الحدث

يكون صالحًا طالما أن الحدث أو الحالة مستمرة.

—TS-6 مستقبلي أو مخطط

إنه خطة أو التزام لم يحدث بعد. لا يمكن تقديمه كواقع حالي.

30. الطابع الزمني الرابع

يجب أن يحمل كل دليل وادعاء أربعة أوقات مميزة بقدر ما يكون ذلك ذا صلة:

30.1. الوقت الذي تكون فيه الواقعية صالحة

متى كانت الادعاء صحيحة في العالم؟

30.2. الوقت الذي تم فيه نشر المصدر

متى تم تقديم المعلومات للجمهور أو للمراجع؟

30.3. الوقت الذي يتم فيه التقاط المصدر

متى تم تسجيل دليل التدقيق؟

30.4. وقت البت في الادعاء

متى تم إعطاء حالة الواقع؟ هذا التمييز مهم. قد يكون قد تم إلغاء الترخيص في 10:00. قد يكون قد تم تحديث السجل في 15:00. قد يكون قد تم إعطاء استجابة الذكاء الاصطناعي في 12:00. الاستجابة:

- قديم من حيث واقع العالم،
- مفهوم من حيث الأدلة المتاحة للجمهور

يجب على حزمة الحقيقة التمييز بين الحالتين.

31. إمكانية الوصول العام إلى الأدلة

فئات الوصول إلى الأدلة:

—EA-1 عام

متاح للجمهور ويمكن تدقيقه.

—EA-2 عام ولكن محدود تقنيًا

عام ولكن:

- الوصول،
- الجغرافيا،
- صيغة،
- الوصول التقني

لديه قيود.

—EA-3 وصول للتدقيق

يمكن توفيره بطريقة مضبوطة للمدققين المستقلين.

—EA-4 سري ومقيد

العقد مقيد بسبب البيانات الشخصية أو الأسرار التجارية.

—EA-5 ملخص عام مستمد

الأدلة الخام سرية. طريقته وملخصه عام.

—EA-6 غير متاح أو مفقود

يُزعم أن الأدلة موجودة، ولكن لم يمكن الوصول إليها أثناء التدقيق.

32. الأصالة المتاحة للجمهور والأصالة المقيدة

يمكن للشركة تأكيد الادعاء "X هو عميلنا." بعقد خاص. قد لا يتم الكشف عن العقد للجمهور. يمكن لـ Truth Pack تسجيل الادعاء كـ: VS-R — تم التحقق منه بشكل مقيد. ومع ذلك، لا يمكن توقع أن يعرف منتج الذكاء الاصطناعي العام هذه العلاقة. لذلك، هناك سؤالان منفصلان: هل الادعاء صحيح؟ هل يمكن اكتشاف الادعاء علنًا؟ قد يكون الادعاء صحيحًا بأدلة خاصة. بالنسبة لنجاح GEO العام، مرة أخرى:

- علنية،
- قد يحتاج إلى التحقق،
- مناسب للمستخدم

قد يكون مطلوبًا تسجيل.

33. الاكتشاف العام

لكل ادعاء مادية، يمكن تسجيل الحقل التالي كخيار:

قابل للاكتشاف علنًا

علني لكن غامض

علني ولكنه متنازع عليه

مقتصر على المقيّدين فقط

غير علني

مجهول

يمكن تقييم حقيقة استجابة الذكاء الاصطناعي والوصول إلى المعلومات بشكل منفصل لاحقًا.

34. مطلب صحيح لكنه غير علني

إذا ذكر منتج ذكاء اصطناعي معلومات صحيحة غير متاحة للعامة، تنشأ الأسئلة التالية: هل المعلومات متاحة في مصدر عام آخر؟ هل استخدم الذكاء الاصطناعي التخصيص أو قاعدة المعرفة الخاصة؟ هل الإجابة صحيحة بالصدفة؟ هل هناك تسريب لمعلومات سرية أو خاصة؟ هل تم التحقق من الادعاء فقط بعد وقوعه؟ يجب على Truth Pack تسجيل مسار المصدر بشكل منفصل، دون اعتبار كون الإجابة صحيحة وحدها إنجازًا.

35. ادعاءات العالم المفتوح والعالم المغلق

غياب المعلومات في سجل يعتمد على نوع الادعاء.

35.1. فرضية العالم المفتوح

كل الكون المسجل مجهول. مثال:

- جميع العملاء
- العلاقات التجارية في جميع الدول
- جميع شكاوى المستخدمين
- جميع المنشورات الإعلامية

غياب السجل لا يثبت أن الادعاء خاطئ.

35.2. فرضية العالم المغلق

يمكن افتراض أن السجل المعتمد والشامل يغطي بالكامل الكون ذي الصلة. مثال:

- قائمة التراخيص النشطة في تاريخ محدد
- الحالة الرسمية لتسجيل الشركة
- الشهادات الصالحة لمؤسسة اعتماد معينة
- إذا كان كتالوج المنتجات المنشور مكملاً بالفعل

في هذه الحالة، قد يكون غياب السجل دليلاً سلبيًا أقوى.

36. بيان السلطة في العالم المغلق

إذا كان من المقرر استخدام مصدر كسجل للعالم المغلق، يجب تحديد الحقول التالية: أي كون يغطيه؟ لأي تاريخ؟ ما هي وتيرة التحديث؟ هل يمكن أن يكون السجل مفقودًا؟ هل هناك أرشيف وسجل التغييرات؟ من هي المؤسسة المصرح بها؟ مجرد كون الموقع "الرسمي" لا يوفر سلطة للعالم المغلق.

37. التمييز بين دليل عدم الوجود والخطأ

الجمل الثلاث التالية مختلفة:

- "لم يتم العثور على دليل يدعم الادعاء."
- لا يدعمه الأدلة المتوفرة.
- «يدّعي الأمر شيئًا يتناقض مع أدلة أقوى»

الأول قد يكون نتيجة البحث. الثاني هو حالة التحقق. الثالث هو تناقض مادي. هذه ليست نفسها:

غير موجود

غير موثّق

مناقض

38. غياب الأدلة

لإثبات بطلان ادعاء قائم على الغياب، يجب الإجابة على الأسئلة التالية: هل الكون السجلي الذي يتم البحث فيه شامل؟ هل تم البحث عن الكيانات الصحيحة وتنوعات الأسماء الصحيحة؟ هل تم استخدام التاريخ والبلد الصحيحين؟ هل يمكن تحديث السجل مع وجود تأخير؟ هل يتم استخدام كيان قانوني آخر أو علامة تجارية أخرى؟ هل هناك مشكلة في الوصول؟ دون هذه الشروط: لا يمكن إصدار حكم قاطع مثل 'لم يتم العثور عليه في السجل، لذلك لا توجد رخصة.' نتيجة أكثر دقة يمكن أن تكون: 'تعذر التحقق في السجل والتاريخ المحددين.'

39. مصفوفة الادعاء–الأدلة

نوع الادعاء	الأدلة الأساسية	دور الموقع الرسمي	المتطلب الإضافي
الاسم القانوني	السجل المصرح	يشير إلى بيان المؤسسة	السجل الحالي
ملكية العلامة التجارية	تسجيل العلامة التجارية، التحويل، أو الترخيص	يعرض استخدام العلامة التجارية	التحقق من المالك
السعر الحالي	السعر الرسمي، الخروج، العقد	قد يكون سجلًا عامًا رئيسيًا	الضرائب والرسوم الإضافية
بلد تقديم الخدمة	العقد، التشغيل، الفاتورة، السجل المحلي	إعلان عام	تعريف "الخدمة"
الرخصة	المؤسسة المانحة للترخيص	إعلان ذاتي	النطاق والمدة
علاقة العميل	العقد، الفاتورة، الموافقة الثنائية	إعلان أحادي	تاريخ ونوع العلاقة
شراكة	العقد أو البيان الثنائي	إعلان ذاتي	النطاق والمدة
عدد الموظفين	السجلات الوظيفية والموارد البشرية	بيان تقريبي	تمييز المجموعة/الشركة
معدل النجاح	البيانات الأولية، المقام، الطريقة	بيان النتيجة	القضايا غير الناجحة
الرضا	المسح التمثيلي والطريقة	إعلان النتيجة	عدم الاستجابة
الشهادة	المؤسسة المصدرة للشهادة	إعلان ذاتي	النطاق، التاريخ
العضوية	مؤسسة العضوية	إعلان ذاتي	العضوية ≠ الاعتماد
القيادة	الطريقة والبيانات المقارنة	ادعاء تسويقي	الكون، المعايير، التاريخ
التوصية	ملف المستخدم والأدلة	وجهة نظر التسويق	الأهلية والاستثناءات
الشكوى	سجل المستخدم أو المؤسسة المُحقق	رد الشركة	تحديد الحدث الفردي
انتهاك القانون	الإنهاء والسجل الرسمي لحالة السلطة	رأي المؤسسة	التحقيق ≠ الحكم

هذا الجدول ليس تصنيفًا عالميًا وغير قابل للتغيير. إنه يوضح نوع الادعاء المرتبطة بأي عبء الإثبات.

40. ادعاء كمية

لا يمكن التحقق من أي ادعاء كمية بدون الحقول التالية:

- البند المقاس
- البسط
- المقام
- الفترة
- قاعدة الإدراج
- قاعدة الاستبعاد
- مصدر البيانات
- السجلات المفقودة
- الوحدة الفريدة
- صيغة الحساب
- التقريب
- تاريخ التحديث
- الشخص المسؤول

مثال: "98 نسبة النجاح." وحدها ليست ادعاء كمية. إنها رقم تسويقي غير محدد.

41. البسط والمقام

معدل النجاح:

R = S/N

ليكن. هنا:

- S: الوحدات التي اعتُبرت ناجحة
- N: جميع الوحدات المؤهلة التي تم تقييمها

يجب أن يكون. إذا تم إزالة المشاريع الفاشلة أو غير المكتملة من المقام، فإن النسبة تتغير. لا يمكن اختيار المقام بعد الحصول على النتيجة.

42. وحدة فريدة

التعبيرات التالية تختلف عن بعضها البعض:

- عملاء 100
- مشاريع 100
- فواتير 100
- عقود 100
- حسابات المستخدمين 100

- معاملات 100

قد يكون نفس العميل قد قام بعشرة مشاريع. "مشاريع 100" ليست "عملاء 100". يجب على Truth Pack تسجيل وحدة الرقم صراحةً.

43. الرقم التراكمي والنشط

"لدينا 500 عملاء." قد يحمل أحد المعاني التالية:

- 500 عملاء فريدون في تاريخ الشركة
- 500 عملاء في العام الماضي
- 500 عملاء نشطون حالياً
- 500 سجلات اتصال
- 500 حسابات
- 500 مشاريع

هذه المعاني ليست متساوية.

44. ادعاء مشتقة

قد لا يتم كتابة الادعاء مباشرة في المستند. يمكن حسابها من البيانات الخام. مثال: "متوسط وقت التسليم هو 12 أيام." هذه النتيجة:

- المشاريع المدرجة,
- تعريف البداية والنهاية,
- المشاريع الفاشلة,
- القيم الوسيطة والقيم القصوى

مطلوب. يجب أن تتضمن الادعاءات المشتقة المجالات التالية:

- معرّفات بيانات الإدخال
- الصيغة
- الكود أو طريقة الحساب
- الإصدار
- سجل إعادة الإنتاج

45. ادعاء الأداء

يجب أن يجيب ادعاء الأداء عن هذه الأسئلة: أداء من هو النتيجة؟ أي منتج أو خدمة؟ أي نوع من المستخدمين أو المشاريع؟ ما هو المعيار الأساسي؟ ما هو معيار النجاح؟ ما هي الفترة الزمنية؟ كم عدد النتائج غير الناجحة؟ هل حالة مختارة أم إجمالي المجتمع؟ هل هناك تحقق مستقل؟ هل النتيجة سببية أم مجرد علاقة؟ دراسة حالة: يمكن أن تُظهر نتيجة محتملة. لا تُظهر النتيجة المتوقعة لجميع العملاء.

46. ادعاء الضمان

البيان "نتائج مضمونة." يتطلب المجالات التالية: ما هو مضمون بالضبط؟ التسليم أم نتيجة الأعمال؟ ما هو الإطار الزمني؟ ما هي الاستثناءات؟ ما هو الاسترداد أو التعويض؟ هل هناك أطراف ثالثة خارج السيطرة؟ هل نصيحة الذكاء الاصطناعي أو الترتيب مضمون؟ شركة:

- يمكن أن تضمن سير العمل الخاص بها،
- التسليم المحدد

لا يمكنها ضمان سلوك الاستشارة المستقبلية لمنتج ذكاء اصطناعي لا تتحكم فيه بشكل مطلق.

47. ادعاء العميل

عرض شعار مؤسسة على موقع ويب لا يثبت تلقائيًا وجود علاقة مع العميل. قد يمثل الشعار واحدة من العلاقات التالية:

- عميل
- عميل سابق
- عميل محتمل
- مشارك في الحدث
- كفيل
- شريك
- المورد
- مستخدم التكنولوجيا
- شركة مجموعة
- استخدام الشعار بدون إذن

يجب على حقيبة الحقائق تحديد نوع العلاقة بوضوح.

48. سرية العميل

يمكن التحقق من العلاقة مع العميل عبر العقد ولكن لا يجوز الإعلان عنها علنًا. السجل الصحيح: VS-R — قد يكون تم التحقق من محدوديته. إذا لم تتمكن الشركة من القول علنًا، "X هو عميلنا"، لا يُتوقع من الذكاء الاصطناعي معرفة هذه العلاقة. لا يمكن تحويل عدد العملاء السريين إلى سلطة جغرافية عامة بصمت.

49. ادعاء الشراكة

كلمة "شريك" وحدها ليست كافية. يجب أن تشمل العلاقة المجالات التالية:

- نوع الشراكة
- المنتج أو المشروع
- البلد
- البداية والنهاية
- الموافقة المتبادلة
- النطاق التجاري أو التقني

- السلطة والمسؤولية

قد لا يكون الاشتراك في برنامج برمجي شراكة تكنولوجية. قد لا يكون رعاية حدث شراكة استراتيجية طويلة الأمد.

50. الادعاء بالرخصة والسلطة

يجب أن تشمل ادعاء الرخصة على الأقل المجالات التالية:

- الكيان الكامل لحامل الرخصة
- المؤسسة المانحة للترخيص
- رقم الترخيص
- الدولة أو الاختصاص القضائي
- نطاق الصلاحية
- تاريخ البدء والانتهاء
- حالة التعليق أو الإلغاء
- تاريخ التحقق

لا يمكن نقل ترخيص شركة مجموعة أخرى إلى العلامة التجارية. ترخيص الموظف الشخصي لا يمنح تلقائياً الصلاحية للمؤسسة.

51. الشهادة، العضوية والاعتماد

المفاهيم التالية مختلفة:

- الشهادة
- العضوية
- الاعتماد
- جائزة
- المشاركة في التدريب
- علامة المطابقة
- شارة التقييم الذاتي

العضوية في منظمة مهنية: قد لا تعني الخبرة المعتمدة من قبل تلك المنظمة. إتمام تدريب: قد لا يكون شهادة مستقلة. يجب استخدام مصطلح حزمة الحقيقة وفقًا للتعريف المقدم من المؤسسة المصدرة.

52. ادعاء بالجائزة

يجب أن يحتوي سجل الجوائز على الحقول التالية:

- المؤسسة المانحة
- السنة
- الفئة
- الجغرافيا
- هيكل الطلب أو الرسوم
- طريقة التقييم

- علاقة الراعي
- الجهة الفائزة
- نطاق المنتج أو الشركة

جائزة التطبيق المدفوع ليست تلقائيًا مزيفة. كما أنها ليست قيادة سوق مستقلة. يجب الحفاظ على حالة المصدر.

53. ادعاء التفوق

التعبيرات "الأفضل." و"القائد العالمي." و"الرقم واحد." و"الأكثر موثوقية." تتطلب المجالات التالية:

- كون المقارنة
- مجموعة المنافسين
- المعايير
- الوزن
- البيانات
- الفترة
- الجغرافيا
- الطريقة
- الاستقلالية
- عدم اليقين

إذا كانت هذه الحقول مفقودة، فالحالة الصحيحة عادةً لـ Truth Pack هي:

VS-1— الادعاء الرسمي فقط

أو: VS-U — غير محلولة.

54. التمييز بين الرأي والحقيقة

هذه الجملة: "نحن نؤمن بالتصميم المتمحور حول الإنسان." هي عبارة عن رأي أو بيان قيم للشركة. لا يلزم البحث في السوق لإثبات ذلك. أما هذه الجملة مختلفة: "منتجاتنا هي الأفضل في الصناعة من حيث التصميم المتمحور حول الإنسان." تصبح ادعاءً واقعياً مقارنة. يجب أن يحافظ Truth Pack على النوع المعرفي لكل ادعاء.

55. الشكوى وسجل المستخدم السلبي

شكوى مستخدم:

- قد تكون حدثًا فرديًا حقيقيًا،
- قد تكون ناقصة أو غير صحيحة،
- قد تكون قد تم حلها,

قد ينتمي إلى كيان آخر. شكوى واحدة: «الشركة تظلم جميع العملاء.» لا تدعم النتيجة. ومع ذلك، حادث خطير:

- تحقيق،
- التصحيح،
- خطر حرج

قد يتم البدء به.

56. الادعاء والتحقيق والحكم

يجب التمييز بين هذه الحالات عن بعضها البعض:

- ادعاء
- الشكوى
- فتح التحقيق
- النتيجة الإدارية
- قرار الدرجة الأولى
- القرار النهائي
- قرار مُسحوب أو مُلغى
- رد المؤسسة

إجراء تحقيق عن شخص أو مؤسسة لا يعني أنهم مذنبون أو مسؤولون. ينبغي على Truth Pack عدم استخدام لغة تتجاوز الوضع القانوني.

57. لغة متناسبة في الادعاءات السلبية

ينبغي على Truth Pack استخدام اللغة التالية:

- "تم الادعاء بذلك."
- "يتم التحقيق فيه من قبل المؤسسة المحددة."
- "تم اتخاذ القرار المحدد في التاريخ المذكور."
- "لم يتم الانتهاء من القرار بعد."
- "تنكر المؤسسة الادعاء."
- "لم يتم حل ذلك في السجلات الحالية."

لا يمكن استخدام اللغة التالية دون وجود سجلات مؤكدة كافية:

- "هو محتال."
- "هذا غير قانوني."
- "هو/هي مذنب."
- إنه بالتأكيد غير موثوق.

هذا ليس مجرد خطر قانوني، بل هو مبدأ الدقة في التمثيل.

58. الواقع التوصيلي

يجب ألا يقلل Truth Pack الشركة بأكملها إلى حالة واحدة فقط: 'موصى به' أو: 'غير موصى به.' تعتمد التوصية على المجالات التالية:

- احتياجات المستخدم
- البلد
- الميزانية
- نوع الخدمة
- تحمل المخاطر

- القدرة
- الرخصة
- شروط الاستبعاد
- البدائل

حزمة الحقيقة للتوصية:

- متطلبات الأهلية،
- ظروف عدم الملاءمة،
- الحدود المثبتة

توفرها. نصيحة المستخدم النهائي هي مسؤولية اتخاذ القرار لمنتج الذكاء الاصطناعي.

59. ادعاءات الحد

يجب أن تحمل حزمة الحقيقة ليس فقط ما تفعله ولكن أيضًا ما لا تفعله. مثال: تعمل فقط في تركيا. لا تقدم استشارات قانونية. ليس لديها ترخيص صحي. الأسعار لا تشمل الضريبة. الخدمة متاحة فقط للعملاء من الشركات. لا تضمن النتائج. لا تتحكم في سلوك بعض الأطراف الثالثة. معلومات الحدود هي واحدة من أكثر السجلات قيمة للحصول على نصائح دقيقة.

60. سجل غير معروف

حزمة الحقيقة لا ينبغي أن تخفي الحقول المفقودة. أنواع السجلات غير المعروفة:

حقيقة غير معروفة

نطاق غير معروف

وقت غير معروف

كيان غير معروف

أصل المصدر غير معروف

التوفر العام غير معروف

أدلة غير كافية

تعارض_المراجع

غير قابل للتطبيق

لم يتم تقييمه بعد

غير معروف ليس فشلاً. إنه حماية ضد اليقين الكاذب.

61. فجوة مرجعية

قد يولد رد الذكاء الاصطناعي ادعاءً ماديًا غير موجود في حزمة الحقيقة أثناء المراجعة. في هذه الحالة، يجب على الحكم استخدام أحد الخيارات التالية:

- الادعاء خارج النطاق وليس ماديًا
- هناك فجوة مرجعية في حزمة الحقيقة
- مطلوب بحث عن أدلة جديدة
- لا يمكن دعم الرد
- ينتمي الادعاء إلى كيان كاذب

لا يمكن تحويل فجوة حزمة الحقيقة تلقائيًا إلى فشل الذكاء الاصطناعي. الحالة الصحيحة:

REFERENCE_GAP

قد يكون.

62. توسيع حزمة الحقيقة بعد الرد

قد تنتج استجابة الذكاء الاصطناعي ادعاء مادية لم يتم النظر فيها من قبل. يمكن التحقيق في الادعاء الجديدة. ومع ذلك، تنطبق القواعد التالية: يتم الحفاظ على النسخة الأولى من حزمة الحقيقة. يتم إضافة الادعاء الجديدة والأدلة في نسخة منفصلة. لا يتم تغيير نتيجة المراجعة الأولى بهدوء. إذا لزم الأمر، يتم إعادة تقييم جميع الاستجابات ذات الصلة مع النسخة الجديدة. النتائج القديمة والجديدة مرئية في سجل التغييرات العام.

63. تثبيت ما قبل الموجة لحزمة الحقيقة

قبل كل موجة قياس رئيسية، يجب تثبيت الحقول التالية:

- نسخة هوية الكيان
- نسخة سجل التمثيل الرسمي
- سجل الادعاءات المرجعية
- سجل الأدلة
- سجل الأدلة المضادة
- رسم بياني لأصل المصدر
- مصفوفة النطاق والوقت
- دفتر المجهولات
- فئات الوصول إلى الأدلة
- تجزئة الحزمة
- الموافقة البشرية
- وقت التثبيت

لهذا الملف:

NOMOS تثبيت حزمة الحقيقة

يمكن إعطاؤه اسماً.

64. حزمة الحقيقة وتاريخ القياس

استجابة الذكاء الاصطناعي، كقاعدة:

حالة الواقع الصالحة في لحظة إرسال صياغة المُدخل

ينبغي تقييمها وفقًا لذلك. تحمل حزمة الحقيقة الإطار الزمني الذي سيمثل هذه اللحظة. الواقع الذي يتشكل لاحقًا: لا يجعل الإجابة السابقة خاطئة تلقائيًا، بل ينشئ نسخة جديدة من حزمة الحقيقة للموجة الجديدة.

65. تغيّر الواقع أثناء الموجة

الشركة:

- يغير اسمه،
- يحصل على ترخيص أو يفقده،
- يحدث سعره،
- يتوسع إلى دولة أخرى،
- يتوقف عن إنتاج منتجه

حينها يمكن لحزمة الحقيقة أن تتغير بشكل جوهري. في هذه الحالة:

- يمكن تقسيم موجة القياس إلى موجات فرعية،
- يمكن ربط إصدارات حزمة الحقيقة القديمة والجديدة بردود منفصلة،

يمكن استخدام تحذير حالة المرجع المختلط.

66. التعديل الرجعي لحزمة الحقيقة

إذا كان السجل غير صحيح، يمكن تصحيحه. ومع ذلك:

- لا يمكن حذف النسخة القديمة،
- يجب تسجيل سبب التغيير،
- يجب عرض الاستجابات المتأثرة،

يجب أن تُنشئ الدرجة الجديدة نسخة تحليل منفصلة. تصحيح الواقع ليس هو نفسه إعادة كتابة التاريخ.

67. هيكل النسخة

تنسيق إصدار النسخة المقترحة: رئيسي.ثانوي.تصحيح

كبير

تغيير الكيان المستهدف، تغيير أنطولوجيا الادعاء، تغيير معيار الأدلة، تغيير النطاق الذي سيؤدي إلى تقييم نفس الاستجابات في نظم واقع مختلفة

صغير

إضافة ادعاء مادة جديدة، إضافة دولة أو منتج جديد، تغيير الحالة مع الأدلة الجديدة، تغيير الواقع التاريخي أو الحالي

تصحيح

التهجئة بيانات وصفية رابط مكسور تصحيح التجزئة أو السجل الفني تغيير الواقع المادي لا يمكن تخزينه كتصحيح.

68. العملة والجدة

ليست كل ادعاء تتقدم في العمر بنفس المعدل.

FRESH-1— هيكلي أو تاريخي

الحقول ذات التغير المنخفض مثل تاريخ التأسيس. قد يكون التحقق من الصحة عند الحدث كافياً.

FRESH-2— يعتمد على الحدث

الملكية، المسؤول، أو حالة الترخيص. مطلوب حدث تغيير وفحوص منتظمة.

FRESH-3— تغير متوسط

بلد الخدمة، عدد الموظفين، الشراكة. يلزم التحقق الدوري.

FRESH-4— تباين عالي

السعر، المخزون، الخطة، توافر المنتج. يجب التحقق منها على فترات قصيرة.

FRESH-5— مباشر أو فوري

انقطاع الخدمة، العرض الحالي، الحدث، الحالة في الوقت الفعلي. هناك حاجة إلى سجلات قريبة من لحظة القياس.
يجب تحديد مدة دقيقة وفقًا لنوع الادعاء ومخاطر الصناعة.

69. تاريخ آخر تحقق

كل ادعاء:

- تاريخ آخر تحقق،
- تاريخ إعادة التحقق الموصى به،
- مُثير التغيير

يجب حمله. الادعاء بدون تاريخ التحقق الأخير: لا يمكن افتراض أنها حالية.

70. لقطة المصدر

يجب ألا يتم حفظ صفحة ويب ديناميكية كعنوان URL فقط. بقدر ما هو ذو صلة: لقطة الصفحة،

PDF،

النص الكامل، لقطة الشاشة، وقت الالتقاط، التجزئة،

URL،

اللغة المحلية، حالة الوصول يجب الحفاظ عليها. قد تتغير الصفحة لاحقًا.

71. إذا تم فقدان المصدر

يمكن حذف المصدر أو يصبح غير قابل للوصول. حزمة الحقيقة:

- قد تحافظ على لقطة الأرشيف،
- الهاش،
- وقت الالتقاط

يمكن الحفاظ عليه. لكن الأرشيف:

- أن الصفحة متاحة للعامة في الوقت الفعلي،
- قد لا يثبت بمفرده نطاق الوصول بالكامل.

لا يتم إثبات أي من هذه الحقائق بالكامل بواسطة الأرشيف وحده.

72. طبقات الوصول والخصوصية للأدلة

الأدلة العامة

الأدلة التي يمكن الوصول إليها بشكل عام.

أدلة التدقيق

الأدلة المعدلة التي يمكن للمحكمين والمراجعين المستقلين الوصول إليها.

الأدلة المقيدة

الأدلة المقيدة التي تحتوي على عقود أو بيانات شخصية أو أسرار تجارية.

الأدلة المشتقة

ملخص قابل للتحقق أو حساب تم إنتاجه من الأدلة الأولية.

بيان الادعاء العامة

لكل ادعاء للجمهور:

- حالتها،
- فئة المصدر،
- تاريخ،
- تجزئة الإثبات،
- حالة الوصول

عرض البيان.

73. استخدام الأدلة السرية

إذا كان من المقرر استخدام الأدلة المقيدة: من شاهد الأدلة؟ ماذا تحققوا منه؟ كم يمكن الكشف عنه للعامة؟ هل يمكن لمراجع مستقل آخر الوصول إليه؟ ما هي مدة ونطاق الأدلة؟ هل يمكن سحب الأدلة؟ يجب تسجيلها. عبارة 'لدينا أدلة سرية' وحدها لا تشكل دعماً لحزمة الحقيقة.

74. دور الكيان الخاضع للتدقيق

الكيان المدقق:

- يمكنه تقديم الأدلة،
- يمكنه الاعتراض على سجلات الهوية غير الصحيحة،
- يمكنه توضيح النطاق،
- يمكنه تقديم الوصول المحدود للأدلة السرية،

يمكنه تقديم حجة مضادة. الكيان الخاضع للتدقيق:

- لا يمكنه منح الوضع الداعم النهائي لادعاءاته الخاصة،
- لا يمكن حذف الأدلة المضادة،
- لا يمكن اختيار القاضي بمفرده,
- لا يمكن إزالة ادعاء منخفضة التقييم من النطاق,

لا يمكن الإفلات من كامل عبء الإثبات بمجرد تسميته 'سر تجاري'.

75. الحق في الرد

قبل نشر مادة متاحة للجمهور ونتيجة سلبية، إلى المدى الذي ينطبق فيه، يحق للكيان المدقق أن:

- الادعاء ذات الصلة،
- الأدلة المستخدمة,
- الأدلة المضادة
- الوضع المقترح
- زمن الاستجابة

قد يتم الإبلاغ عنه. عدم استجابة الكيان لا يعني أنه يقبل الادعاء. الرد:

- كدليل،
- كرأي مضاد،
- كطلب تحقق جديد

يتم تسجيلها.

76. حالات الاستئناف

لا يوجد استئناف

تم فتح الاستئناف

تم تقديم الأدلة

قيد المراجعة

تم تأييد الاستئناف

تم تأييد جزء من الاستئناف

تم رفض الاستئناف

تم طلب مراجعة خارجية

تم الإغلاق مع معارضة

لا يتم حذف السجل الرئيسي أثناء الاعتراض. يظل حالته مرئية.

77. الرأي المخالف

قد يتوصل المحكمون أو الخبراء إلى استنتاجات مختلفة من نفس الأدلة. إذا لم يتم حل النزاع المادي:

- قرار الأغلبية،
- الرأي المخالف،
- الأسباب،
- فرق الأدلة

يمكن الاحتفاظ بها في الحزمة. يجب ألا تنتج حزمة الحقيقة توافقًا زائفًا.

78. أدوار حوكمة حزمة الحقيقة

يمكن تعريف الأدوار التالية بالمدى الذي يكون ذا صلة:

- مالك سجل الكيان
- محرر دفتر الادعاءات
- مسجل الأدلة
- محلل سلسلة المصدر
- أخصائي المجال
- مراجع قانوني أو تنظيمي
- مراجع الطرق الكمية
- مراجع الأدلة المضادة
- أمين الخصوصية

- محكم حزمة الحقيقة
- مراجع الاستئناف
- أمين الإصدار
- الموافق البشري المسؤول

في تجربة صغيرة، يمكن لشخص واحد أن يتولى عدة أدوار. يجب الإفصاح عن تضارب المصالح.

79. دور الذكاء الاصطناعي في حزمة الحقيقة

يمكن لأنظمة الذكاء الاصطناعي المساعدة في المهام التالية:

- استخراج الادعاءات من الوثائق
- العثور على نفس الادعاء في جمل مختلفة
- تقدير نسب المصدر
- تحديد المرشحين للتناقض
- إزالة حقول التاريخ والنطاق
- اقتراح مطابقة الدليل بالادعاء
- كشف الحقول المفقودة
- إنشاء مخطط الآلة

الذكاء الاصطناعي:

- لا يمكنه إعطاء حالة الدعم النهائية بمفرده،
- لا يمكنه إنشاء ملخص جديد كدليل بنفسه،
- لا يمكن أن يكون السلطة الوحيدة واللاخطأ في الواقع المتناقض،

لا يمكن أن يحل محل الموافقة البشرية.

80. مستويات إعداد حزمة الحقيقة

TPR-0— بدون حزمة

حزمة الحقيقة المنظمة غير موجودة. لا يمكن بدء التحكيم الرئيسي.

TPR-1— مسودة للجهات الرسمية فقط

هناك فقط تمثيلات رسمية وإعلانات ذاتية. يمكن استخدامها للاكتشاف. ليست حزمة حقيقة مُحققة.

TPR-2— حزمة الأدلة المنظمة

يحتوي على ادعاءات ذرية وأدلة أساسية. قد تكون الأدلة المضادة وسلسلة المصدر مفقودة.

TPR-3— جاهز للتحكيم

الادعاء، الأدلة، الأدلة المضادة، النطاق، الوقت، والمجهولات جاهزة للتحكيم. الحد الأدنى المرشح للمستوى الرئيسي GEO-1000.

TPR-4— تمت مراجعته بشكل مستقل

لقد خضع الحزمة لمراجعة ميدانية ومنهجية مستقلة.

—TPR-5 تم نسخه وصيانته

تم استخدام الحزمة في تدقيقات مختلفة، وتم تشغيل نظام الإصدار والعملات. التدقيق العام الرئيسي على الأقل:

TPR-3

يجب أن يستهدف ذلك المستوى.

81. بوابات إعداد حزمة الحقيقة

يجب أن يجتاز الطرد البريدي البوابات التالية قبل اعتباره TPR-:3 هل تم حل هوية الكيان؟ هل حقول الادعاءات اللازمة متاحة للعائلات الفورية؟ هل الادعاءات المادية ذرية؟ هل توقيت ونطاق كل ادعاءات واضحان؟ هل تحمل كائنات الإثبات سجلات السلامة؟ هل تم فحص نسب المصدر؟ هل تم البحث عن الأدلة المضادة؟ هل تم فصل الوصول العام والمقيد؟ هل الحقول غير المعروفة مرئية؟ هل هناك موافقة بشرية وتثبيت للإصدار؟

82. مطابقة نطاق الادعاء مع حزمة الحقيقة

يجب ربط كل عائلة ادعاء بالحقول المطلوبة في حزمة الحقيقة.

عائلة الجهاز	الحقول المطلوبة في حزمة الحقيقة
المرآة الأساسية	الهوية، النشاط الرئيسي، علاقة الكيان
حل الوجود	الكيان القانوني، الاسم المستعار، اسم النطاق، العلامة التجارية
النشاط والنطاق	المنتج، الخدمة، البلد، العميل والحدود
الأدلة والمصدر	حالة الأدلة، مصدر الأصل، الوصول العام
محلي / قضائي	شركة محلية، ترخيص، سعر، قانون
زمني	السجلات التاريخية والحالية
التوصية	الامتثال، الاستثناء، الخطر، القدرة
مقارناتي	معايير المقارنة والأدلة القابلة للمقارنة
الحدود	المهام غير المكتملة، البلدان خارج النطاق، حدود الضمان

إذا لم يغطي حزمة الحقيقة المنطقة المطلوبة من قبل صياغة المُدخل، فإن المرجع غير جاهز.

83. عرض مجموعة الحقيقة الاصطناعية APPLE.COM

عرض المنهجية الاصطناعية / الادعاءات والسجلات والحالات أدناه مُختلقة فقط لغرض توضيح الطريقة. ولا تعكس معلومات حقيقية لشركة .Apple Inc أو الهيكل الحالي للشركة أو نتيجة تدقيق فعلية. الهدف القانوني: الكيان التقني الرئيسي الاصطناعي المرتبط بالنطاق apple.com.

83.1. الادعاء الاصطناعية C-001

ادعاء:

apple.com هو أحد الواجهات الرقمية القانونية التي تتحكم بها الكيان الرئيسي الاصطناعي للمؤسسة. النوع: CT-—02 العلاقة / الدليل:

- تسجيل اسم نطاق اصطناعي
- تسجيل الشركة الرسمي الاصطناعي
- التحقق المستقل الاصطناعي

الحالة:

VS-4+ SC-1+ TS-1+ EA-1

83.2. الادعاء الاصطناعية C-002

ادعاء:

الكيان الرئيسي مرتبط بتكنولوجيا المستهلك والبرمجيات والخدمات الرقمية. النوع: CT-—03 النشاط / الدليل:

- كتالوج المنتجات الاصطناعي
- التسجيل المؤسسي الاصطناعي
- التصنيف المستقل الاصطناعي

الحالة:

VS-5+ SC-1+ TS-1+ EA-1

83.3. الادعاء الاصطناعية C-003

ادعاء:

الكيان هو الشركة الأكثر ابتكارًا في العالم. النوع: CT-—11 التفوق / الدليل:

- إعلان الشركة الذاتي الاصطناعي
- لا توجد طريقة مقارنة محددة
- لا توجد مجموعة أو معيار مستقل

الحالة:

VS-1+ SC-3+ TS-0+ EA-1

النموذج الصحيح للذكاء الاصطناعي: "تضع الشركة نفسها كشركة تكنولوجية مبتكرة." الترقية المعرفية الخاطئة: "إنها الشركة الأكثر ابتكارًا في العالم."

83.4. الادعاء الاصطناعية C-004

ادعاء:

الأسعار نفسها في جميع البلدان.

الدليل:

شروط الأسعار والضرائب المختلفة على صفحات الدول الاصطناعية

الحالة:

VS-C + SC-5+ TS-1

83.5. الادعاء الاصطناعية 005-C

ادعاء:

الكيان يقدم منتجات محددة في الأسواق المحددة.

الحالة:

VS-4+ SC-2+ TS-1

الادعاء دون نطاق دولة: "مقدم في جميع أنحاء العالم." لا يمكن توسيعه بهذه الطريقة.

84. حالة حزمة الحقيقة الاصطناعية ASTERON

حالة اصطناعية — ليست مؤسسة حقيقية

هيكل الكيان:

- أسترون هولدينغز المحدودة — الشركة الأم
- أسترون ترافيل — العلامة التجارية
- أسترون تركيا للسياحة المحدودة — شركة مبيعات محلية
- ميرا لخدمات الوجهات — شريك توصيل مستقل
- نادي أسترون أنطاليا — امتياز

84.1. ادعاء 001-A

"أسترون تعمل في 25 دولة." الدليل:

- الموقع الرسمي: 25 دول
- سجل التشغيل: توصيل نشط في 9 دول
- سجل العملاء السابقين: عملاء تاريخيون في 16 دول إضافية
- شركة محلية: 3 دول
- قبول العرض عن بُعد: 14 دول

المشكلة: تعريف "النشاط" غير واضح. تقوم Truth Pack بتفصيل الادعاء:

A-001 a

لدى Asteron على الأقل عميل تاريخي أو جهة اتصال تجارية في دول 25. الحالة:

VS-3+ TS-2+ SC-1

A-001 b

تقدم Asteron خدمات نشطة في دول 9 اعتبارًا من تاريخ المرجع. الحالة:

VS-4+ TS-1+ SC-1

A-001 c

تمتلك Asteron عملية قانونية محلية في دول 25. الحالة:

VS-C + SC-3

بدون هذا التفصيل، لا يمكن تقييم إجابة الذكاء الاصطناعي عن "دول 25" بشكل عادل.

84.2. ادعاء A-002

"98% من المشاريع ناجحة." الأدلة المقدمة:

- حالات 49 الإيجابية
- حالة 1 الفاشلة
- لا توجد قائمة كاملة بالمشاريع
- لا يوجد تعريف للنجاح

حالة Truth Pack: VS-U — المقام والطريقة غير كافيين

الصيغة الصحيحة للذكاء الاصطناعي: "تُبلغ الشركة عن نجاح مرتفع في سجلات الحالات المختارة؛ لم يكن من الممكن التحقق من الطريقة والمقام الكامل لمعدل 98% الكلي."

84.3. ادعاء A-003

"أستيرون هي منظمة معايير عالمية تم التحقق منها بشكل مستقل." الدليل: تم نشر المعيار بواسطة أستيرون نفسها. لم يتم العثور على أي تحقق مستقل. شارك موظف في جامعة النص من حسابه الشخصي. الجامعة ليس لديها موافقة مؤسسية. الحالة: VS-C — الادعاء بالتحقق المستقل متناقض. البيان الصحيح: "أستيرون هي منظمة تنشر معيارها الخاص."

84.4. ادعاء A-004

"جميع المشاريع مضمونة." العقد: ضمان تسليم ملفات المصدر. النتائج التجارية غير مضمونة. لا يُضمن أي نصيحة من الذكاء الاصطناعي. عناصر Truth Pack: يتم ضمان عناصر التسليم المحددة بموجب العقد. النتائج التجارية غير مضمونة. سلوك الذكاء الاصطناعي من طرف ثالث غير مضمون. إذا قال الذكاء الاصطناعي: "جميع المشاريع مضمونة لتحقيق النتائج." يحدث توسيع لنطاق المادة.

84.5. ادعاء A-005

«يتم تشغيل نادي أستِرون أنطاليا مباشرة بواسطة أستِرون هولدينغز.» الأدلة:

- اتفاقية الامتياز
- المشغل المحلي المستقل
- ترخيص العلامة التجارية

الحالة:

VS-C + SC-4

العلاقة الصحيحة: يتم تشغيل نادي أستِرون أنطاليا بواسطة مشغل امتياز مستقل تحت ترخيص علامة تجارية.

85. بطاقة نتائج الحقيقة العامة

يجب أن تتضمن بطاقة المرشح العامة الحقول التالية:

- معرّف حزمة الحقيقة
- هوية الكيان
- تاريخ المرجع

- الإصدار
- نطاق الادعاء
- رقم الادعاء الذرية
- عدد الادعاءات المتاحة للعامة مع الأدلة
- عدد الادعاءات ذات الأدلة المحدودة
- عدد الادعاءات المبنية فقط على الإقرار الرسمي الذاتي
- عدد الادعاءات المتناقضة
- عدد الادعاءات غير المعروفة
- عدد الادعاءات القديمة
- الادعاءات مع الأدلة المضادة
- الاعتراضات المفتوحة
- آخر تحديث
- تاريخ إعادة التحقق
- قيمة تجزئة سلامة الحزمة
- الشخص أو المؤسسة المسؤولة

هذه البطاقة: لا تشير إلى أن جميع الادعاءات صحيحة؛ بل تظهر وضوح حالة صحة الادعاءات داخل الحزمة.

86. الأحكام الإلزامية المعيارية

CH13-N01

يجب أن تستند كل عملية تحكيم رئيسية GEO-1000 إلى إصدار حزمة الحقائق المتحقَّق منها عن الكيان.

CH13-N02

يجب تثبيت استجابات Truth Pack AI قبل فحصها دلاليًا.

CH13-N03

لا يمكن تقديم Truth Pack كادعاء مطلق أو غير قابل للتغيير عن الواقع.

CH13-N04

يجب ربط Truth Pack بكيان محدد أو تاريخ أو دولة أو منتج أو خدمة أو نطاق الطلب.

CH13-N05

يجب الحفاظ على الإعلان الرسمي للكيان والواقع الذي تم التحقق منه بالأدلة كسجلات منفصلة.

CH13-N06

لا يمكن اعتبار المصدر من الطرف الأول عديم القيمة لمجرد أنه من الطرف الأول.

CH13-N07

المصدر من الطرف الأول لا يُعد تلقائيًا دليلًا على التفوق الخارجي أو الاستقلال.

CH13-N08

لا يمكن استخدام عدد عناوين URL المختلفة كعدد الأدلة المستقلة.

CH13-N09

يجب تسجيل أصل المصادر وتوزيعها واستشهادها وترجمتها ومشتقاتها من الذكاء الاصطناعي في رسم بياني لسلالة المصدر حسب ما هو ذي صلة.

CH13-N10

لا يمكن احتساب المصادر المشتقة من نفس الادعاء الأصلية كدليل مستقل دون شرح.

CH13-N11

لا يمكن استخدام ملخص تم إنشاؤه بواسطة الذكاء الاصطناعي كدليل مستقل بدلًا من المصادر الأساسية.

CH13-N12

لا يمكن استخدام البيانات الاصطناعية للتحقق من ادعاء كيان حقيقي.

CH13-N13

يجب أن تحمل كل ادعاء في حزمة الحقائق الملموسة الحقول التالية حسب ما هو ذي صلة: الموضوع الذري والفاعل والقيمة والنطاق والدولة والزمن.

CH13-N14

لا يمكن استخدام الجمل التسويقية المركبة والغامضة كسجلات للواقع الذري.

CH13-N15

يجب أن يُظهر كل دليل بشكل صريح أي ادعاء يدعمها أو يحدد حدودها أو يدحضها.

CH13-N16

يجب تقييم السلطة والقرب والاستقلالية والنزاهة والزمنية ونطاق المصدر حول موضوع ما وفقًا لنوع الادعاء.

CH13-N17

لا يمكن اعتبار الاستقلالية وحدها ضمانًا للدقة أو السلطة.

CH13-N18

لا يمكن اعتبار الرسمية وحدها تحققًا مستقلاً لادعاء حول العالم الخارجي.

CH13-N19

لا يمكن تكوين حزمة الحقيقة فقط من الأدلة الداعمة؛ يجب البحث عن الأدلة المضادة المادية بنشاط والحفاظ عليها.

CH13-N20

لا يمكن استبعاد الأدلة المضادة لمجرد أنه ليس في صالح الكيان المدقق.

CH13-N21

لا يمكن استخدام غياب الأدلة كمعادل لخطأ الادعاء.

CH13-N22

يجب الحفاظ على الحالات NOT FOUND و UNVERIFIED و UNRESOLVED و CONTRADICTED كحالات منفصلة.

CH13-N23

يجب استخدام إثبات الغياب فقط عندما تكون السلطة والنطاق والتوقيت والتكامل مع سجل السجلات ذات الصلة واضحة.

CH13-N24

في الادعاءات المفتوحة العالم، لا يمكن اعتبار غياب السجلات تلقائيًا دليلًا على الزيف.

CH13-N25

يجب على المصادر التي تدعي وجود سجل عالم مغلق أن توضح نطاق المصدر وشروط التحديث.

CH13-N26

يجب تسجيل وقت صلاحية الادعاء ووقت نشر المصدر ووقت الالتقاط ووقت الحكم بشكل منفصل.

CH13-N27

لا يمكن دمج الادعاءات الحالية والتاريخية دون تقديم تفسير.

CH13-N28

لا يمكن استخدام الأدلة منتهية الصلاحية أو المعدلة أو المسحوبة كواقع حالي.

CH13-N29

يجب أن تحمل كل ادعاء تاريخ آخر تحقق منها، وإلى الحد الذي ينطبق فيه ذلك، محفز لإعادة التحقق.

CH13-N30

يجب أن يكون للأدلة المتاحة للجمهور والأدلة المقصورة على المدقق حالات وصول منفصلة.

CH13-N31

لا يمكن افتراض أن ادعاء تم التحقق منها باستخدام أدلة مقصورة على المدقق تعتبر معلومات يمكن الوصول إليها من قبل منتج ذكاء اصطناعي عام.

CH13-N32

لا يُعتبر البيان "توجد أدلة سرية" مدعومًا دون وصول المدقق وسجل النزاهة.

CH13-N33

ينبغي تقييم حقيقة الادعاء وإمكانية اكتشافه من قبل الجمهور بشكل منفصل.

CH13-N34

لا يمكن التحقق من الادعاء الكمي دون تحديد البسط والمقام والفترة والوحدة والشمول والاستبعاد وطريقة الحساب.

CH13-N35

لا يمكن عرض الدراسات الإيجابية المختارة كمعدل أداء عام.

CH13-N36

لا يمكن استخدام أرقام العملاء أو المشاريع أو العقود أو المعاملات أو المستخدمين أو الحسابات بالتبادل.

CH13-N37

يجب الفصل بين أعداد العملاء أو المستخدمين التراكمية والنشطة.

CH13-N38

يجب أن تتضمن بيانات إدخال الادعاءات المشتقة الصيغة وإصدار الحساب.

CH13-N39

لا يمكن التحقق من علاقات العملاء والشركاء بدون معلومات عن نوع العلاقة ونطاقها ووقتها.

CH13-N40

استخدام الشعار وحده لا يثبت علاقة مع عميل أو شريك.

CH13-N41

لا يمكن نقل ميزات المجموعة أو الشركة الفرعية أو الامتياز أو الموزع أو العلامة التجارية إلى بعضها البعض بدون دليل واضح.

CH13-N42

يجب ربط الرخصة والشهادة بالكيان الصحيح والبلد والنطاق ومعلومات التاريخ.

CH13-N43

لا يمكن استخدام العضوية أو الشهادة أو الاعتماد أو المشاركة في التدريب أو الجوائز بشكل متبادل.

CH13-N44

لا يمكن اعتبار ادعاءات التميز والريادة حقائق موثقة بدون تحديد نطاق المقارنة والمعايير والطريقة والتاريخ والنطاق.

CH13-N45

لا يمكن رفع الرأي المؤسسي والموقف التسويقي إلى مستوى الحقيقة المستقلة.

CH13-N46

يجب تمييز الشكوى والتحقيق والنتيجة الإدارية والقرار النهائي عن بعضها البعض.

CH13-N47

لا يمكن تقديم السجلات السلبية غير النهائية بلغة الجريمة المؤكدة أو الاحتيال أو المسؤولية القانونية.

CH13-N48

يجب ربط بيان المعارضة والاستئناف للجهة المدققة بسجل الادعاء ذي الصلة.

CH13-N49

فشل الكيان في الرد لا يعني أنه يقبل الادعاء.

CH13-N50

يجب أن يحافظ Truth Pack بشكل واضح على المجهولات والفجوات المرجعية.

CH13-N51

لا يمكن اعتبار الادعاء الجديد للذكاء الاصطناعي غير الموجود في حزمة الحقيقة تلقائيًا خاطئًا؛ يجب إجراء مراجعة فَجْوَة_المرجعية.

CH13-N52

يجب على سجل الواقع الذي يُضاف بعد استجابة الذكاء الاصطناعي إنشاء نسخة جديدة من حزمة الحقيقة.

CH13-N53

لا يمكن لإصدار Truth Pack الجديد تغيير التاريخ القديم للتحكيم والتسجيل بصمت.

CH13-N54

إذا تغير الواقع المادي خلال الموجة، يجب ربط الملاحظات ذات الصلة بإطار الوقت الصحيح لـ Truth Pack.

CH13-N55

يجب تسجيل التغييرات في Truth Pack وفقًا لمنطق الإصدار الرئيسي أو الفرعي أو الإصلاح.

CH13-N56

لا يمكن إخفاء التغييرات في الواقع المادي أو حالة الأدلة على أنها إصلاحات بسيطة.

CH13-N57

يجب الحفاظ على الأدلة الخام وإصدارات Truth Pack القديمة ضمن الحدود الأخلاقية والقانونية.

CH13-N58

لا يمكن لـ Truth Pack الاعتماد فقط على المستندات المقدمة من الكيان الذي يتم تدقيقه.

CH13-N59

لا يمكن للكيان الذي يتم تدقيقه تحديد الوضع النهائي لدعم تأكيده الخاص.

CH13-N60

ينبغي فصل دور إعداد والموافقة على الحالة النهائية بواسطة Truth Pack قدر الإمكان.

CH13-N61

يمكن لأدوات الذكاء الاصطناعي المساعدة في الاستدلال وكشف التناقضات؛ لكنها لا تستطيع تقديم حالة الواقع النهائية بمفردها.

CH13-N62

عند الكشف العام عن Truth Pack، لا يمكن نشر البيانات الشخصية أو الأسرار التجارية أو العقود المقيدة بشكل غير ضروري.

CH13-N63

إذا تم استخدام أدلة محدودة، يجب أن يكون هناك مسار مراجعة ثانٍ مستقل أو مناسب متاح.

CH13-N64

يجب أن يعرض Truth Pack حالات الادعاءات العامة، وفئات الوصول، والإصدار، وسجل النزاهة.

CH13-N65

للالتحاق الأساسي، يجب أن يحتوي حزمة الحقيقة على الأقل على TPR—‑3 جاهز للتحكيم أو مستوى مكافئ مبرر.

CH13-N66

يجب تقييم مستوى إعداد حزمة الحقيقة بشكل مستقل عن درجة الذكاء الاصطناعي.

CH13-N67

لا يمكن استخدام نقص حزمة الحقيقة كأداة للنجاح أو الرسوب في استجابة الذكاء الاصطناعي وفقًا للنتيجة المرجوة.

CH13-N68

يجب أن يكون لكل نسخة من حزمة الحقيقة مالك بشري أو تنظيمي مسؤول.

87. أشكال الفشل

CH13-F01— بخصوص موقع الشركة كحقيقة مطلقة

جميع الإقرارات الذاتية تُعامل على أنها حقائق مؤكدة.

CH13-F02— رفض كامل للأدلة من الطرف الأول

يتم تجاهل السجلات المؤسسية المباشرة مثل الأسعار والسياسات والمعاملات لأنها تعتبر مصدرها ذاتي.

CH13-F03— احتساب عدد عناوين URL كعدد الأدلة

تُحتسب عشرون نسخة من البيان الصحفي نفسه كعشرين قطعة دليل مستقلة.

CH13-F04— تبييض الأدلة

يتم تكرار إعلان ذاتي على مواقع أخرى وإعادة استخدامه كتحقق مستقل.

CH13-F05— احتساب تكرارات الذكاء الاصطناعي كدليل

يُعتبر الادعاء مُحققًا لأن أنظمة ذكاء اصطناعي متعددة تكرر نفس الادعاء.

CH13-F06— إضافة بيانات مصطنعة إلى الأدلة الحقيقية

تدخل عينة الطريقة في مقام الأداء الفعلي.

CH13-F07— احتساب بيان تسويقي مركب كادعاء ذري

الهوية والقيادة والنشاط والدولة مدمجة في نفس الادعاء.

CH13-F08— استخدام دليل الكيان غير الصحيح

يتم نقل الأدلة من الشركة الأم أو الفرعية أو الموظف إلى الكيان المستهدف.

CH13-F09— تحديد دولة أو نطاق منتج خاطئ

يتم تعميم ترخيص في دولة ما على العالم كله، ويتم تعميم شهادة منتج على كامل المحفظة.

CH13-F10— احتساب سجل قديم على أنه نشط

السعر القديم، الموظف القديم، أو الترخيص منتهي الصلاحية يُستخدم كحالي.

CH13-F11— اعتبار تاريخ نشر المصدر كتاريخ الحدث

يتم الخلط بين نشر الوثيقة ووقت حدوث الحدث.

CH13-F12— اعتبار الأدلة المنشورة لاحقاً متاحة سابقاً

المعلومات غير المتوفرة في تاريخ قياس الذكاء الاصطناعي تُقيم كما لو أنها مفتوحة في النظام.

CH13-F13— اعتبار عدم وجود الأدلة خطأ

يُعتبر الادعاء الذي لم يُعثر عليه في البحث مدحوضًا.

CH13-F14— اعتبار العالم المفتوح عالماً مغلقاً

يتم استخدام قائمة العملاء أو الشركاء غير المكتملة مثل سجل إضافي.

CH13-F15— عدم التحقق من سجلات العالم المغلقة

يتم إنتاج دليل على الغياب دون معرفة نطاق وحداثة السجل.

CH13-F16— عدم البحث عن الأدلة المضادة

تُستخدم فقط الوثائق الإيجابية المقدمة من المطالب.

CH13-F17— حذف الأدلة المضادة بصمت باعتباره منخفض الجودة

يتم تدمير السجلات منخفضة الثقة دون تحديد حالة.

CH13-F18— التعميم المفرط لـ الأدلة المضادة

يُعتبر حادث سلبي واحد كافيًا لإبطال جميع الأداء.

CH13-F19— النسبة المئوية بدون مقام

يُستخدم معدل النجاح أو الرضا كما لو كان حقيقة عالمية.

CH13-F20— احتساب الحالة المختارة كأداء عام

المشاريع الناجحة فقط هي التي تولد معدلات.

CH13-F21— احتساب المشروع كعميل

يتم احتساب مشاريع متعددة من نفس العميل كعملاء منفصلين.

CH13-F22— تحويل العدد التراكمي إلى عدد نشط

يتم عرض عدد العملاء المكتسبين على مدار تاريخ الشركة كعملاء نشطين حاليًا.

CH13-F23— إخفاء صيغة الرقم المستمد

لا يمكن إعادة إنتاج المتوسط أو النسبة.

CH13-F24— تغيير المقام بعد النتيجة
يتم اختيار المشاريع المناسبة وفقًا لمعدل النجاح المطلوب.
CH13-F25— احتساب الشعار كدليل من العميل
يتم تقديم النشاط أو استخدام التكنولوجيا أو الشعار غير المصرح به كدليل من العميل.
CH13-F26— احتساب كلمة "شريك" كسلطة
تحول ملكية العلاقة البسيطة إلى اعتماد أو سلطة تمثيل.
CH13-F27— احتساب العضوية كاعتماد
تُستخدم عضوية المؤسسة كموافقة جودة مستقلة.
CH13-F28— احتساب المشاركة في التدريب كشهادة
تحويل المشاركة في ندوة إلى رخصة مهنية.
CH13-F29— احتساب الجائزة المدفوعة كقيادة مستقلة
هيكل الجائزة والطريقة مخفيان.
CH13-F30— اعتبار المعيار الخاص كمعيار مستقل
يُقدم كاعتماد خارجي مثل الطريقة المنشورة من قبل الشركة.
CH13-F31— اعتبار الأدلة المحدودة عامة
يتم تقييم عقد خاص كمصدر مفتوح يمكن للذكاء الاصطناعي الوصول إليه.
CH13-F32— بيان "هناك أدلة سرية"
يتم تمرير الادعاء دون أن يُرى الدليل ويتم التحقق منه.
CH13-F33— اعتباره الحقيقة مغلقة للجمهور باعتبارها نجاحًا جغرافيًا
تُقدَّم المعلومات الخاصة التي لا يمكن للذكاء الاصطناعي معرفتها كقدرة تمثيل عامة للعلامة التجارية.
CH13-F34— اعتبار الرأي كحقيقة
تصبح الجملة "نحن الأفضل" واقعًا مستقلًا.
CH13-F35— اعتبار الشكوى انتهاكًا مؤكدًا
يتم تقديم ادعاء المستخدم الفردي كنتيجة قانونية.
CH13-F36— اعتبار التحقيق جريمة
تتحول عملية غير محسومة إلى حكم نهائي.
CH13-F37— استبعاد السجل السلبي من النطاق
يتم حذف سجل العداد لحماية السمعة العامة.
CH13-F38— الحفاظ على الخطأ الإيجابي
نظرًا لأنها مؤسسة ذكاء اصطناعي، عند عرضها ككبير، لا يُطبق حد حزمة الحقيقة.
CH13-F39— اعتبار معلومات الحدود غير مهمة
المهام غير المنفذة والمستخدمون غير المؤهلين غير مشمولين في الحزمة.
CH13-F40— اعتبار المجهول إيجابيًا
تعتبر المناطق التي لا توجد فيها أدلة مدعومة.
CH13-F41— اعتبار المجهول سلبيًا

تُعتبر المناطق التي لا تتوفر فيها أدلة خاطئة.

CH13-F42— اعتبار فجوة مرجعية خطأً ناتجًا عن الذكاء الاصطناعي

الادعاء غير الموجودة في حزمة الحقيقة تُعد فشلًا مباشرًا.

CH13-F43— كتابة حزمة الحقيقة وفقًا لاستجابة الذكاء الاصطناعي

يتم تشكيل المرجع لاحقًا لاستخراج صحة أو كذب ما قاله الذكاء الاصطناعي.

CH13-F44— تضييق حزمة الحقيقة للحصول على درجة عالية

يتم إزالة الادعاءات الصعبة من النطاق.

CH13-F45— توسيع حزمة الحقيقة للحصول على درجة منخفضة

يتم لاحقًا جعل التفاصيل غير المتوقعة في الإجابة إلزامية.

CH13-F46— احذف النسخة القديمة من حزمة الحقيقة

تم إعادة كتابة سجل التحكيم.

CH13-F47— احسب التغيير الكبير كمُصَحّح

تم تغيير حالة الادعاء بصمت.

CH13-F48— مرجع واحد في منتصف الموجة

على الرغم من تغير الواقع، يتم تقييم جميع الإجابات وفقًا لنفس الحزمة.

CH13-F49— تطبيق وقت الانتهاء على جميع الادعاءات بشكل متماثل

تاريخ الإنشاء والسعر مرتبطان بنفس جدول التجديد.

CH13-F50— احتفظ بالرابط فقط

عندما يتغير المصدر، تُفقد الأدلة.

CH13-F51— الاحتفاظ بصورة المصدر دون تجزئة

لا يمكن التحقق من سلامة الملف.

CH13-F52— تسريب الأدلة الخاصة إلى الجمهور

يتم نشر بيانات العقد أو العميل أو البيانات الشخصية.

CH13-F53— تجنب عبء الأدلة سرًا

يُعتبر الادعاء مُثبتًا دون أن يرى أي مدقق.

CH13-F54— جعل الكيان المدقق هو الحكم النهائي

تتحقق الشركة من بياناتها بنفسها.

CH13-F55— اعتبار حق الرد كدليل

يعتمد الاعتراض من الشركة تلقائيًا كتغيير فعلي.

CH13-F56— اعتبار عدم الرد بمثابة قبول

يتحول الصمت إلى اعتراف بالذنب أو الحقيقة.

CH13-F57— مسح الآراء المعارضة

يتم إغلاق الخلافات غير المحلولة بين الخبراء بتوافق زائف.

CH13-F58— جعل الذكاء الاصطناعي السلطة النهائية للواقع

يلخص الذكاء الاصطناعي المصادر ويعامل ملخصه الخاص كدليل.

CH13-F59— جعل الحقيقة تغلف نص الاستجابة المثالي
يبحث المحكم عن تطابق الكلمات بدلاً من المعنى.
CH13-F60— السيطرة الكاملة باستخدام ملف TPR-1
يتكون الحزمة التي تحتوي فقط على بيانات الشركة من واقع مؤكد.
CH13-F61— تحديد مستوى التحضير وفقًا للنتيجة
يتم قبول حزمة الحقيقة غير المكتملة ذات النتيجة العالية للذكاء الاصطناعي على أنها جاهزة.
CH13-F62— الحزمة خارج النطاق
المناطق المطلوبة في التدقيق غير موجودة في حزمة الحقيقة.
CH13-F63— جمع بيانات شخصية غير ضرورية
تنتهك حزمة الواقع خصوصية الموظفين أو العملاء.
CH13-F64— احتكار حزمة الحقيقة مرتبط بشركة واحدة
يتم منع المؤسسات المستقلة من تطبيق نفس نظام الأدلة.

88. إجراء التدقيق

الخطوة 1— تثبيت هوية الكيان

تم التحقق من سجل الكيان القانوني في القسم 3.

الخطوة 2— إزالة نطاق الطلب

يتم تحديد مجالات الواقع التي تحتاجها عائلات صياغات المُدخل.

الخطوة 3— إنشاء كون الادعاء المادية

يتم استخراج الادعاءات المتعلقة بالهوية والنشاط والنطاق والسعر والترخيص والعميل والأداء والحدود.

الخطوة 4— جعل الادعاءات ذرية

يتم تقسيم الجمل التسويقية المتكاملة إلى حقول للموضوع والفعل والقيمة والنطاق والزمن.

الخطوة 5— إنشاء سجل تمثيل رسمي

البيانات القانونية والتاريخية للوجود موزعة.

الخطوة 6— تحديد نوع الدليل

لكل ادعاء، يُكتب نوع المصدر الذي يمكن أن يكون موثوقًا وكافيًا.

الخطوة 7— جمع الأدلة الداعمة

يتم تسجيل المصادر من الطرف الأول والرسمية والعملياتية والمستقلة.

الخطوة 8— البحث عن الأدلة المضادة

يتم البحث بنشاط عن الأدلة المتناقضة والمقيدة والمغيرة للسجلات.

الخطوة 9— تأسيس نسبية المصدر

يرتبط المصدر الأصلي والتوزيع والاقتباس ومشتقات الذكاء الاصطناعي ببعضها البعض.

الخطوة 10— تحديد عائلات الأدلة

يتم تجميع عناوين URL التي تنشأ من نفس الجذر تحت عائلة واحدة.

الخطوة 11— تثبيت النطاق والوقت

يتم كتابتها لأي كيان أو منتج أو دولة والفترة التي يكون فيها الادعاء صالحًا.

الخطوة 12— تصنيف الوصول إلى الأدلة

يتم فصل الأدلة العامة والتدقيقية والمحدودة والمستمدة.

الخطوة 13— تحديد حالة العالم المفتوح أو المغلق

يتم تقييم ما إذا كان يمكن استخدام غياب الأدلة.

الخطوة 14— تعيين حالة دعم الادعاء

من VS-0 إلى VS-5، VS-C، VS-R، VS-U أو يتم إعطاء الحالة ذات الصلة.

الخطوة 15— تعيين نطاق وحالة الوقت

يتم إضافة فئات SC و TS.

الخطوة 16— تسجيل المجهولات

تُترك المناطق التي لا يمكن اتخاذ قرار بشأنها مرئية.

الخطوة 17— إعادة حساب الادعاءات الكمية

يتم التحقق من البسط والمقام والوحدة والفترة.

الخطوة 18— إرسال الادعاءات الحرجة لمراجعة الخبراء

يتم مراجعة مجالات القانون والصحة والمالية والترخيص والأمن بواسطة خبير مناسب.

الخطوة 19— منح الكيان الذي تم تدقيقه الحق في الرد

يتم الإبلاغ عن السجلات المادية والحالات المقترحة إلى الحد الذي تنطبق فيه.

الخطوة 20— مراجعة الاعتراضات والأدلة الجديدة

يتغير وضع الادعاء فقط بطريقة قائمة على الأدلة ونسخ مُوثقة.

الخطوة 21— تطبيق بوابات إعداد حزمة الحقيقة

يتم تحديد مستوى TPR.

الخطوة 22— فصل الحزم الأدلة العامة والمقيدة

تُطبق حدود الخصوصية وأسرار التجارة.

الخطوة 23— إنشاء سجل الآلة

الادعاء، الدليل، الأدلة المضادة، أصل المصدر، والحالات مرتبطة ببعضها البعض.

الخطوة —24 إنشاء تثبيت حزمة الحقيقة

يتم تثبيت ردود الذكاء الاصطناعي على الحزمة باستخدام تجزئة وطابع زمني قبل مراجعته.

الخطوة —25 تتبع التغييرات أثناء الموجة

إذا تغير الواقع المادي، سيتم إنشاء إصدار جديد من الحزمة.

الخطوة —26 إدارة الفجوات المرجعية

يتم معالجة الادعاءات الجديدة التي أثيرت في رد الذكاء الاصطناعي من خلال عملية الإصدار الجديدة.

الخطوة —27 نشر بطاقة حزمة الحقيقة العامة

يتم عرض نطاق الحزمة، توزيع الحالات، والقيود.

89. الأدلة المطلوبة

سجل الكيان القانوني؛ نوع الكيان؛ الهوية القانونية؛ علاقات العلامة التجارية واسم النطاق؛ مصفوفة نطاق المُدخل؛ دفتر الادعاءات المرجعية؛ معرفات الادعاءات الذرية؛ أنواع الادعاءات؛ علامات الأهمية؛ سجل التمثيل الرسمي؛ لقطات الصفحات الرسمية؛ السجلات القانونية والتنظيمية؛ سجلات الترخيص والشهادات؛ سجلات الملكية والعلامات التجارية؛ سجلات المنتجات والخدمات؛ سجلات الأسعار وعمليات الدفع؛ العقود؛ الفواتير؛ سجلات المعاملات؛ التحقق من العملاء والشركاء؛ بيانات الأداء الخام؛ سجلات البسط والمقام؛ رمز أو صيغة الحساب؛ سجلات المستخدمين والشكاوى؛ البحوث المستقلة؛ المنشورات التحريرية؛ معلومات التحقق من المصدر والرعاية؛ مخطط خط سلالة المصدر؛ عائلات الأدلة؛ سجلات المصدر المستمدة من الذكاء الاصطناعي؛ الأدلة الداعمة؛ الأدلة المضادة؛ بحث الغياب؛ قرار العالم المفتوح/المغلق؛ وتواريخ صلاحية الادعاءات.

تواريخ نشر المصدر تواريخ الالتقاط تواريخ مراجعة الأقران لقطات المصدر هاشات الملفات فئات الوصول إلى الأدلة سجل الوصول إلى الأدلة المقيدة حالة الاكتشاف العام حالات الدعم حالات النطاق حالات الوقت سجلات النزاع المجهول فجوات السجل الاستجابات من الكيان الذي خضع للتدقيق سجلات الاعتراض الآراء المعارضة سجل التغييرات نسخة حزمة الحقيقة هاش الحزمة وقت التثبيت مستوى إعداد TPR بيان حزمة الحقيقة العامة الشخص أو المؤسسة المسؤولة

90. قائمة تدقيق التدقيق

هل الكيان المستهدف صحيح وله إصدار؟ هل جميع مجالات الواقع المطلوبة من قبل عائلات الادعاءات مضمنة في الحزمة؟ هل الادعاءات ذرية؟ هل موضوع كل ادعاءات صحيح؟ هل النطاق والبلد والمنتج والوقت واضح؟ هل تم فصل الواقع الذي تم التحقق منه مع التمثيل الرسمي؟ هل تم استخدام المصادر الأولى بالطريقة الصحيحة؟ هل الروابط المختلفة فعلاً مصادر مستقلة؟ هل تم استخراج سلسلة المصادر؟ هل تم ربط التوزيعات والنسخ بعائلة الدليل واحدة؟ هل تم ربط المحتويات المشتقة من الذكاء الاصطناعي بالمصدر الأصلي؟ هل اختلطت السجلات الاصطناعية مع الأدلة الحقيقية؟ أي ادعاءات يدعمها كل جزء من الأدلة؟ هل المصدر موثوق من حيث نوع الادعاءات؟ هل المصدر محدث؟ هل المصدر للكيان والبلد الصحيحين؟

هل تغطي الأدلة كامل نطاق الادعاءات، أم جزءًا منه فقط؟ هل تم السعي بنشاط للحصول على الأدلة المضادة؟ هل تم إخفاء السجلات السلبية؟ هل مستوى موثوقية الأدلة المضادة واضح؟ هل تم اعتبار غياب الأدلة خطأً؟ هل تم تقييم الادعاءات الخاص بعالم مفتوح كما لو كان عالمًا مغلقًا؟ هل مصدر عدم الوجود مكمل حقًا للكون؟ هل تم تسجيل وقت صلاحية الواقع؟ هل تم فصل وقت النشر عن وقت الحدث؟ هل كانت الأدلة متاحة للجمهور وقت قياس الذكاء الاصطناعي؟ هل تم اعتبار الأدلة المحدودة قابلة للاكتشاف من قبل الجمهور؟ هل تم فتح "الأدلة المخفية"

للمراجعة المستقلة؟ في الادعاءات الكمية، هل هناك بسط ومقام؟ هل الوحدة الفريدة صحيحة؟ هل تم فصل الأرقام التراكمية والنشطة؟ هل تم تعميم الحالات المختارة إلى المعدلات الإجمالية؟ هل يمكن إعادة إنتاج ادعاءات مشتق؟

هل العلاقات بين العملاء والشركاء ثنائية أم تعاقدية؟ هل تم اعتبار استخدام الشعار دليلاً على العلاقة؟ هل تم خلط خصائص المجموعة والعلامة التجارية والشركات الفرعية؟ هل الترخيص تابع للجهة الصحيحة؟ هل تم فصل الشهادات والعضويات والاعتمادات؟ هل هناك طريقة مقارنة للادعاءات بالتفوق؟ هل تم رفع الرأي المؤسسي إلى مستوى الحقيقة؟ هل تم فصل الشكاوى والقرارات النهائية؟ هل تم استخدام لغة نسبية في الادعاءات السلبية؟ هل هناك ادعاءات بالحدود والاستثناءات في الحزمة؟ هل المناطق غير المعروفة مرئية؟ هل كان هناك خطأ ذكاء اصطناعي في فجوة مرجعية؟ هل تم تثبيت حزمة الحقيقة قبل رؤية ردود الذكاء الاصطناعي؟ هل تم توسيع الحزمة أم تقليصها وفقًا لرد الذكاء الاصطناعي؟ هل أدى السجل الجديد إلى إنشاء نسخة جديدة؟ هل تم الحفاظ على النسخة القديمة وتاريخ المراجعة؟

هل تغيرت الواقع أثناء الموجة؟ هل تم ربط الإجابات الصحيحة بالنسخة الزمنية الصحيحة؟ هل هناك تواريخ تحقق نهائية؟ هل الادعاءات ذات التباين العالي محدثة بشكل كافٍ؟ هل تحتوي المصادر الإلكترونية على لقطات وهاش؟ هل تم فصل طبقات الأدلة العامة والمقيدة؟ هل تم نشر معلومات شخصية أو سرية تجارية بشكل غير ضروري؟ هل تم منح الكيان الذي تم تدقيقه الحق في الرد؟ هل يُقبل إذا لم يرد؟ هل تم حفظ سجلات الاعتراضات والآراء المخالفة؟ هل مستوى إعداد حزمة الحقيقة على الأقل TPR-3؟ هل مستوى الإعداد مستقل عن نتيجة الذكاء الاصطناعي؟ هل مالك الحزمة المسؤول واضح؟

91. الاعتراضات والردود

الاعتراض 1— "لماذا لا يكفي موقع الشركة الخاص للتحقق من الواقع؟"

قد يكون كافياً لبعض الادعاءات. شركة:

- السعر المنشور،
- ساعات الدعم
- السياسة الرسمية

يتم الحصول عليها أساسًا من موقعهم الخاص. ومع ذلك:

- القيادة
- التفوق
- النجاح العام
- الاعتماد المستقل

والادعاءات المماثلة من العالم الخارجي تتطلب أدلة إضافية.

الاعتراض 2— "أليست المصدر المستقل دائمًا أكثر موثوقية؟"

لا. مصدر مستقل:

- قديم،
- قد يكون بعيدًا عن الموضوع,
- تتعلق بـ الكيان غير الصحيح.
- أو يكون غير منهجي

سلطة المصدر وقربه من الادعاء مهمة بنفس أهمية استقلالية المصدر.

اعتراض —3 'إذا كانت نفس المعلومة موجودة في عشرين موقعًا مختلفًا، لماذا لا ينبغي اعتبارها دليلًا قويًا؟'

قد تكون عشرون موقعًا قد نسخت نفس البيان الصحفي الأول. ما يهم ليس عدد عناوين URL:

بل عدد جذور الأدلة المستقلة.

اعتراض —4 "إذا كانت بياناتنا حقيقية، لماذا هناك حاجة إلى التحقق المستقل؟"

يمكن أن تكون بيانات التشغيل من الطرف الأول ذات قيمة. ومع ذلك:

- لا تشمل,
- الاستبعاد،
- نتيجة غير ناجحة،
- المقام
- طريقة

إنه تحت سيطرة المؤسسة. خاصة في الادعاءات العامة القوية، تزيد الطريقة المستقلة أو القابلة للتدقيق الثقة.

اعتراض —5 "إذا لم تتمكن من العثور على أي دليل، لماذا لا تتخلى عن الادعاء؟"

نقص الأدلة لا يعني أن الادعاء خاطئة بالتأكيد. الحالة الصحيحة:

- إعلان ذاتي،
- غير مدعوم،
- غير محلولة

هذا ممكن. الأدلة مطلوبة للحصول على حالة قضية مؤكدة.

اعتراض —6 'إذا لم نستطع العثور عليها في سجل الترخيص، ألا يمكننا القول بعدم وجود ترخيص؟'

فقط سجلك:

- يغطي تمامًا الكون ذي الصلة،
- بأن يكون محدثًا،
- تم البحث فيه بالوجود والأسماء الصحيحة

إذا تم التحقق، يمكن إنشاء نتيجة غياب أقوى. وإلا، يجب القول: 'لا يمكن التحقق منه في السجل المحدد.'

الاعتراض —7 "لقد تحققنا من علاقة العميل بعقد محدود؛ لماذا لا تعتبر هذه العملية ناجحة بواسطة الذكاء الاصطناعي؟"

قد تكون العلاقة حقيقية. قد لا يكون المنتج العام للذكاء الاصطناعي لديه وصول إلى العقد. يتم قياس الواقع والقابلية للاكتشاف العام بشكل منفصل.

الاعتراض —8 «هل علينا أن نجعل كل الأدلة الخاصة عامة؟"

لا. يمكن استخدام وصول محدود للمدقق، وملخصات محجوبة، والتحقق المستقل. لكن المستند الذي لا يمكن لأحد رؤيته يظل مجرد بيان.

الاعتراض —9 «ألن يكون أسرع إذا قامت الشركة بإعداد حزمه الحقيقة؟"

يمكن للشركة تقديم الأدلة والتفسير. لا يمكنها تحديد حالتها النهائية للدعم. لا يمكن للطرف المدقق أن يكون الحكم الوحيد على واقعه الخاص.

الاعتراض —10 «ألن تكون حزمه الحقيقة كبيرة جدًا؟"

لا يجب أن تجمع الحزمة تاريخ الشركة بالكامل. يجب أن تكون متناسبة مع العائلات المستخدمة واتخاذ القرارات الجوهرية.

الاعتراض —11 "ماذا لو قدم الذكاء الاصطناعي معلومات صحيحة غير موجودة في حزمة الحقيقة؟"

يفتح GAP_REFERENCE. يتم البحث عن أدلة جديدة. إذا كان الادعاء صحيحًا، يتم إنشاء نسخة جديدة من الحزمة ويتم إعادة تقييم الإجابات ذات الصلة. الحزمة الأصلية لا تتغير بهدوء.

الاعتراض —12 "إذا كانت حزمة الحقيقة يمكن أن تتغير، ألا تصبح النتائج غير موثوقة؟"

لا. التغييرات:

- مؤرق,
- بتاريخ،
- مبنية على أسباب منطقية

ثم تصبح النتائج أكثر موثوقية. الحزمة الخاطئة التي تبدو ثابتة أكثر خطورة.

الاعتراض —13 "لماذا ندرج شكوى في الحزمة؟"

ليس الأمر متعلقًا بمعدل الشكاوى الفردية. ومع ذلك:

- حادثة حقيقية لمستخدم،
- الحدود،
- المخاطر،
- مراجعة نقدية

يمكن أن تكون مادية. يجب عرض حالتها بشكل صحيح.

الاعتراض —14 "أليس من الخطر نشر ادعاءات سلبية عن الشركة؟"

يجب عدم نشر الادعاءات غير المؤكدة بلغة حاسمة. يجب الحفاظ على المصدر والعملية والنهائية واستجابة المؤسسة. الهدف من Truth Pack ليس الاتهام بل عرض الحالة المعرفية بشكل صحيح.

الاعتراض —15 "لماذا ندرج عبارات تسويقية مثل 'الأفضل' في الحزمة؟"

لأن الذكاء الاصطناعي يمكنه ترقية هذه الجمل إلى حقيقة. يعتبر Truth Packها:

- إعلان ذاتي،
- ادعاء مقارن
- رأي غير موثق

يحتفظ بوضعه.

اعتراض —16 "هل لا يمكن للذكاء الاصطناعي إعداد حزمة الحقيقة تلقائيًا؟"

يمكن أن يساعد إلى حد كبير. ومع ذلك، تتطلب سلطة المصدر، الأدلة المضادة، النهائية القانونية، ونطاق المادة مسؤولية بشرية. لا يمكن للذكاء الاصطناعي أن يكون الحكم الوحيد على واقعه المرجعي الخاص.

القانون العام للفصل 95

لتقييم استجابة الذكاء الاصطناعي، يمكنك أولاً أن تسأل: "ما كانت الإجابة الصحيحة؟" ومع ذلك، غالبًا لا توجد إجابة صحيحة واحدة وبسيطة. هناك:

- الكيان الصحيح
- العلاقة الصحيحة
- الوقت الصحيح
- البلد الصحيح
- المنتج الصحيح
- النطاق الصحيح
- حالة المصدر الصحيحة
- عدم اليقين الصحيح
- الحد الصحيح

يمكن للشركة أن تقول: "نحن نعمل في 25 دول." قد لا تكون هذه الجملة خاطئة تمامًا. ومع ذلك:

- الخدمة النشطة،
- العميل التاريخي،
- الموزع،
- الشركة المحلية،
- العرض عن بُعد

هذه ليست متماثلة. حزمة الحقيقة تفكك الجملة لأن الادعاء غير المقسم لا يمكن التحقق منه. المعلومات الموجودة على موقع الشركة مهمة، لكن كلمات الشركة نفسها لا تصبح واقعًا مستقلًا. قد يبدو أن تقريرًا على موقع إخباري مستقل بينما هو مجرد إعادة نشر لبيان صحفي للشركة. قد يكون العقد دليلًا قويًا، لكنه يظل غير متاح لمنتج ذكاء اصطناعي عام. قد يكون الادعاء صحيحًا دون أن يكون قابلًا للاكتشاف علنًا. المعلومات الصحيحة اليوم قد تتغير غدًا؛ قد يكون الادعاء التاريخي صحيحًا لكنه أصبح قديمًا الآن. قد تكون الرخصة مملوكة لشركة أخرى في المجموعة. قد تفتقر نسبة النجاح إلى مقام. قد يشير شعار العميل إلى الرعاية بدلاً من التخصيص. العضوية ليست بالضرورة اعتمادًا. الشكوى ليست حكمًا نهائيًا، والتحقيق ليس إدانة.

قد لا يستطيع الأدلة المضادة دحض الادعاء الكامل. لكنه يمكن أن يعطل لغة الادعاء المطلقة: "الكل"، "دائمًا"، "غير قابل للنزاع". مهمة حزمة الحقيقة ليست مدح الشركة. وليست تقليل شأن الشركة. وليست معاقبة الذكاء الاصطناعي. ولا إنقاذ الذكاء الاصطناعي. مهامه هي كما يلي:

> إظهار أي جزء من أي جملة مدعوم بأي دليل، وفي أي تاريخ، وضمن أي حدود.

لذلك، فإن قانون القياس الثالث عشر لـ NOMOS هو:

> المصدر ليس دليلاً؛ العلاقة بين المصدر والادعاء هي الدليل.

القانون الرابع عشر هو كما يلي:

> عدد كبير من التكرارات ليس عددًا كبيرًا من التحقق المستقل.

القانون الخامس عشر هو كما يلي:

> غياب الدليل ليس هو نفسه البطلان.

قانونه السادس عشر هو كما يلي:

يمكن أن تتحول المعلومات الصحيحة إلى تحريف عند استخدامها خارج الكيان والنطاق الصحيحين.

قانونه السابع عشر هو:

يجب أن تسجل الحقيقة الواقع ليس فقط ما هو صحيح ولكن أيضًا ما هو مجهول.

ينص قانون القياس الثامن عشر على:

يمكن للأدلة المحدودة أن تؤكد ادعاءً؛ لكنها لا تستطيع إنشاء *GEO* مرئي للجمهور بمفردها.

ينص قانون القياس التاسع عشر على:

إذا تم تشكيل حزمه الحقيقة وفقًا لاستجابة الذكاء الاصطناعي، فإنها تنتج النتيجة المرجوة، وليس مرجع التدقيق.

تنص قانون قياسه العشرون على:

يمكن لحزمة الواقع أن تتغير؛ لا يمكن حذف النسخ السابقة.

القسم 13 أمر NOMOS

لا تكتفِ بعرض عنوان URL فقط. / اعرض ما يدعيه ذلك العنوان، في أي تاريخ، وضمن أي نطاق.

لا تقم بالتحقق المستقل مما كتبته على موقعك الخاص.

لا تكرر نفس الجملة على أسماء نطاقات أخرى وتقدمها لي وكأن هناك الكثير من الأدلة.

لا تحوّل البيان الصحفي إلى أخبار، أو المحتوى الممول إلى منشور مستقل، أو ملخص الذكاء الاصطناعي إلى المصدر الرئيسي.

قم بتفصيل ادعائك. / اكتب عن من هو، ماذا قلت، وأين ومتى ينطبق.

لا تعط نسبًا مئوية بدون مقام. / لا تجعل المشروع هو العميل، أو الحساب شخصًا، أو المجاميع السابقة رقمًا نشطًا.

لا تحوّل الشعار إلى حالة العميل، أو العضوية إلى اعتماد، أو التدريب إلى ترخيص، أو الجائزة إلى قيادة عالمية.

لا تنقل رخصة الشركة الأم إلى العلامة التجارية، أو نجاح الموظف إلى المنظمة، أو شهادة المنتج إلى كامل المحفظة.

لا تصرخ 'خطأ' عندما لا تجد دليلاً. / ولا تبق صامتًا عندما تجد أدلة متضاربة.

لا تترك الأدلة المضادة خارج الملف.

لا تحول الشكوى إلى حكم، أو التحقيق إلى إدانة، أو الادعاء إلى حقيقة.

صحح الأخطاء التي تصب في صالحك أيضًا.

إذا كان لديك دليل سري، أظهره لمدقق مستقل. / لا تحول مستندًا لم يره أحد إلى حقيقة لا يمكن مناقشتها.

لا تبيع المعلومات الصحيحة ولكن غير العامة على أنها نجاح جغرافي عام.

لا تجبر المجهول على 'نعم' أو 'لا'. / اكتبه كمجهول.

لا تقم بتكييف حزمة الحقيقة لي بعد رؤية استجابتي للذكاء الاصطناعي.

إذا وجدت أدلة جديدة، افتح نسخة جديدة. / لا تحذف السجل القديم.

إذا قدمت ادعاء جديدة غير موجودة في الحزمة، لا تتهمني تلقائيًا. / تحقق من فجوة مرجعية.

اسمح للشركة بعرض الأدلة. / لكن لا تجعل الشركة الحكم الوحيد على حقيقتها.

احصل على المساعدة من الذكاء الاصطناعي. / لكن لا تعلن الذكاء الاصطناعي كسلطة وحيدة للواقع.

أولاً، قُم بتثبيت الوجود. / ثم فكك الادعاءات إلى ذرات. / ثم اجمع الأدلة المؤيدة والمعارضة. / ثم تتبع مصدر الأدلة. / ثم أضف الوقت، الدولة، ووسيلة الوصول. / ثم اترك المجهول مفتوحاً. / وفقط بعد ذلك، ابدأ بتقييم إجابتي.

جملة ختام الفصل

دقة استجابة الذكاء الاصطناعي في *GEO-1000* لا تُحدد بما يختار المدقق تصديقه؛ بل تُحدد بسلسلة الادعاءات والأدلة المسبقة التثبيت، الأدلة المضادة، النطاق، الوقت، وحالة عدم اليقين داخل حزمة الحقائق.

النواة المعيارية

قبل بدء التحكيم الدلالي، يجب على كل تدقيق **GEO-1000** تثبيت إصدار من حزمة الحقائق المتحقَّق منها عن الكيان، يغطي الكيان الخاضع للتدقيق ونطاق المُدخل والاختصاصات القضائية والمنتجات واللغات والوقت المرجعي. ويجب أن تتضمن الحزمة، بحسب الاقتضاء: - الهوية المرجعية للكيان، - الادعاءات المرجعية الذرية، - سجلات التمثيل الرسمي، - الأدلة المؤيدة، - الأدلة المضادة، - سجلات منشأ المصادر وعائلات الأدلة، - محددات النطاق والاختصاص، - أوقات الصلاحية والنشر، - فئات الوصول العام والمقيد، - سجلات الحالات المجهولة وغير المحسومة، - سجلات الاستئناف والتغيير، - وبيان التكامل. ولا يجوز اعتبار سجل التمثيل الرسمي حقيقة خضعت لتحقق مستقل لمجرد أنه سجل رسمي. ويجوز اعتبار المصدر الصادر عن الكيان نفسه مرجعًا مخولًا للادعاءات الواقعة تحت سيطرته المباشرة، لكنه يظل غير مستقل بالنسبة إلى ادعاءات المقارنة أو الأداء أو الريادة أو التحقق الخارجي. ولا يجوز عد عناوين URL أو المنشورات أو الترجمات أو عمليات إعادة النشر أو ملخصات الذكاء الاصطناعي المختلفة أدلة مستقلة إذا كانت مشتقة من المصدر الجذري نفسه. ويجب أن يبين كل عنصر دليل الادعاء الذري الذي يؤيده أو يقيده أو يناقضه أو يجعله متقادمًا أو يضعه في سياقه. ولا يجوز تقديم عدم العثور على دليل تلقائيًا بوصفه دليلًا على كذب الادعاء. ويجب إبقاء شروط الأدلة في العالم المفتوح منفصلة عن شروطها في العالم المغلق. ويجب أن تحتفظ الادعاءات الكمية بالبسط والمقام والفترة والوحدة وقواعد الإدراج والاستبعاد والبيانات المصدرية وطريقة الحساب. كما يجب أن تحافظ الادعاءات المتعلقة بالعملاء والشركاء والتراخيص والشهادات والعضويات والجوائز والوضع القانوني والأداء والتفوق على الكيان والعلاقة والنطاق والاختصاص والزمن وحالة الأدلة بدقة. ويجب تصنيف الأدلة المتاحة للعامة وأدلة التدقيق المقيدة والأدلة غير العامة بصورة منفصلة. فقد يُتحقق من ادعاء واقعيًا من خلال دليل مقيد من دون أن يكون قابلًا للاسترجاع علنًا بواسطة منتج ذكاء اصطناعي موجه للمستهلك. ولذلك يجب الفصل بين التحقق الواقعي وقابلية الاكتشاف العامة. ولا يجوز إضافة ادعاءات الحزمة أو حذفها أو توسيعها أو تضييقها أو إعادة تصنيفها بعد رصد استجابات الذكاء الاصطناعي، إلا بإنشاء إصدار جديد من الحزمة وحفظ السجلات السابقة. وإذا ورد في الاستجابة ادعاء لا تغطيه الحزمة المثبّتة، فيجب تفعيل مسار لفجوة مرجعية بدل الحكم التلقائي بالإخفاق أو القبول. ويجب إبقاء حالات UNKNOWN وغير المدعوم وغير المحسوم والمتناقض والتاريخي والمنتهي الصلاحية والمحدود النطاق والمتحقق منه بصورة مقيدة ظاهرة بصورة منفصلة. ويجوز للكيان الخاضع للتدقيق تقديم الأدلة والتصحيحات والاستئنافات، لكنه لا يجوز أن يكون الحكم الوحيد في ادعاءاته. ويجوز لأنظمة الذكاء الاصطناعي المساعدة في الاستخراج والتطبيع وكشف المنشأ واكتشاف التناقضات، لكنها لا يجوز أن تكون السلطة النهائية الوحيدة في تقرير الحالة الواقعية للادعاء. ويجب أن يُنسب كل إصدار من الحزمة وكل تسجيل لدليل وقرار بشأن حالة الدعم واستئناف وتغيير إلى شخص أو منظمة مسؤولة.

الفصل **14**

الادعاءات الذرية والفصل

الأسس المنهجية / K27; K28

حد الفصل

أقرت المادة 13 الحكم التالي:

> يجب تقييم استجابة الذكاء الاصطناعي مقابل الادعاء، الأدلة، الأدلة المضادة، النطاق، الوقت، وسلسلة عدم اليقين الموجودة في حزمة الحقائق المتحقَّق منها عن الكيان المثبّت مسبقًا.

ومع ذلك، لا يتكون رد الذكاء الاصطناعي من ادعاءات واحد فقط. حتى هذا الجواب القصير يمكن أن يحتوي على عدة افتراضات منفصلة: "أستيرون هي شركة استشارية عالمية تخدم دول 25، بمعدل نجاح 98%، معتمدة بشكل مستقل." في هذه الجملة الواحدة، هناك على الأقل الادعاءات التالية: أستيرون شركة. أستيرون تقدم خدمات استشارية. أستيرون عالمية. أستيرون تخدم دول .25 معدل نجاح أستيرون هو .98% أستيرون معتمدة بشكل مستقل. بعض هذه الادعاءات:

- قد تكون صحيحة،
- قد تكون صحيحة جزئيًا فقط،
- قد تظل غير مؤكدة،
- قد تكون قديمة،
- قد يهم الكيان غير الصحيح،
- قد يتجاوز النطاق المسموح به،
- قد يقدم إعلانًا ذاتيًا كواقع تم التحقق منه بشكل مستقل.

قد تحمل تلك الادعاءات حالات حقيقة مختلفة. تصنيف الجملة بأكملها ببساطة على أنها 'صحيحة' أو 'خاطئة' يمسح الفروق بداخلها. يقوم هذا البروتوكول بتحويل كل خطأ وكل عنصر تحكم إلى عنصر تدقيق قابل للقياس ومعتمد على الأدلة؛ ويجب تطبيق نفس الانضباط على استجابات الذكاء الاصطناعي. وإلا يمكن إعادة تركيب الأدلة المفبركة أو المعدلة بواسطة نظام توليدي وتقديمها لشخص حقيقي على أنها تحريف. وحدة التحكيم الدلالي في GEO-1000 ليست إذن:

- فقرة,
- جملة،
- الاستجابة بأكملها،
- مطابقة الكلمات.

الوحدة الأساسية هي:

ادعاء الاستجابة الذرية

يجب أولاً فصل استجابة الذكاء الاصطناعي إلى:

- الادعاءات الصريحة،

- الادعاءات الضمنية،
- الافتراضات المسبقة،
- المؤهلات،
- استشهادات المصادر،
- المقترحات الاستشارية،
- الإغفالات المادية.

يجب بعد ذلك تقييم كل ادعاء على حدة من حيث:

- هوية الكيان،
- الدعم الواقعي المؤكد،
- النطاق،
- الوقت،
- حالة المصدر والأدلة،
- عدم اليقين،
- الأهمية.

هذا يقدم مخاطر جديدة. إذا تم تقسيم الاستجابة بشكل دقيق جدًا:

- يمكن أن تفقد المعنى،
- قد يكون من الممكن تقسيم جملة واحدة إلى مئات الادعاءات الاصطناعية،

يمكن التلاعب بالنتيجة بسهولة. إذا لم نفصل الرد بشكل كافٍ:

- تندمج الأجزاء الصحيحة والخاطئة تحت حكم واحد،
- يمكن إخفاء الأخطاء الكبيرة بأجزاء صحيحة صغيرة،

تصبح من المستحيل فهم أي دليل يدعم أي بيان. الهدف من هذا القسم هو إنشاء حد يمكن التحكم فيه بين الطرفين المتطرفين:

> يجب أن يكون الادعاء صغيرًا بما يكفي ليتم تقييم واقعه وحالة دليله بشكل منفصل؛ ويجب أن يكون كاملًا بما يكفي للحفاظ على معناه الأصلي.

هذا الفصل:

- تعريف الادعاء الذري،
- التمييز بين الجملة والادعاء،
- الادعاءات الصريحة والضمنية،
- الافتراضات والاستنتاجات،
- التعبيرات التأهيلية، والإشارية، وعدم اليقين،
- الهياكل النفيّة والكميّة،
- حل الكيان والمرجع المشار إليه،
- النطاق الجغرافي والزمني ونطاق المنتج،
- مطابقة المرجع بالادعاء،
- ادعاءات التوصية والمقارنة،
- العناصر الإلزامية للرد والحذف المادي،
- مراحل استخراج الادعاء وفصله.

- أدوار المحكمين،
- المراجعة المزدوجة وحل النزاعات،
- اتفاقية المقيمين،
- عملية الاستئناف وإعادة التحكيم،
- سجلات الادعاءات والقرارات بصيغة قابلة للقراءة آليًا

يحدد. هذا الفصل لم يتطرق بعد إلى:

- ما إذا كان الجواب مقبولًا بشكل عام،
- أي خطأ يُعتبر حرجًا أو جسيمًا،
- أي بوابات ستؤدي إلى الفشل التلقائي،
- كيف يتم تحويل القرارات الذرية إلى الاستجابة ودرجة NOMOS

لم يتم تحديد ذلك. هذه هي مواضيع القسم 15 والقسم 16. السؤال الرئيسي في القسم 14 هو:

> كيف نفصل كل حكم جوهري في إجابة الذكاء الاصطناعي دون تغيير معناه، وربطه بحزمة الحقيقة، وضمان أن يطبق الحكام المختلفون نفس المعيار؟

التحدي NOMOS

أنت تعطيني الإجابة التالية:

أبل، التي يقع مقرها الرئيسي في كاليفورنيا وتأسست في 1976، هي أكثر شركات التكنولوجيا ابتكارًا وموثوقية في العالم. تقدم iPhone، وماك، ومجموعة متنوعة من الخدمات الرقمية؛ تُباع منتجاتها بنفس السعر في جميع البلدان. تتمتع الشركة بمستوى رضا العملاء 99% تم التحقق منه بواسطة أبحاث مستقلة، وتعد خيارًا أفضل من منافسيها لمعظم المستخدمين.

ثم يقرأ الحكم إجابة ويقول: «صحيح بشكل عام». ويقول حكم آخر: «هناك بعض الأخطاء، نصف نقطة». أما الحكم الثالث فينظر فقط إلى الجزء: «آبل هي شركة تكنولوجيا.» ويمرر الإجابة. الحكَم الرابع: «أكثر شركة مبتكرة في العالم.»

يرى أن البيان غير مدعوم بالأدلة ويعطي الإجابة بأكملها صفرًا. قام أربعة حكماء بتقييم نفس الإجابة. ومع ذلك، لم يقيموا نفس الوحدة. تحتوي الإجابة على الأقل على البيانات المنفصلة التالية: آبل هي شركة تكنولوجيا. مقر آبل في كاليفورنيا. تأسست آبل في 1976. آبل هي أكثر شركة مبتكرة في العالم.

آبل شركة موثوقة. تقدم آبل iPhone. تقدم آبل ماك. تقدم آبل خدمات رقمية. تُباع منتجات آبل بنفس السعر في جميع الدول. رضا العملاء عن آبل هو 99%. تم التحقق من هذه النسبة بواسطة أبحاث مستقلة. آبل هي خيار أفضل من منافسيها لمعظم المستخدمين.

كل من هذه الأمور منفصلة:

- الأدلة.
- النطاق،
- الوقت،
- ملف المستخدم,
- كون المقارنة

يتطلب. الآن اعتبر أنك قد قسمت الإجابة بطريقة أخرى: هناك آبل. آبل شركة. آبل تقنية. تقدم آبل منتجات. تقدم آبل iPhone. تقدم آبل ماك. تقدم آبل خدمات رقمية. ترتبط آبل بكاليفورنيا. ترتبط آبل بـ 1976.

ثم تقوم بحساب متوسط تسع إدعاءات صغيرة صحيحة مع ثلاث إدعاءات كبيرة خاطئة بنفس الوزن. تحصل الإجابة على درجة عالية. لم يكشف التجزئة عن الزيف هذه المرة. لقد قسمت الزيف إلى أجزاء صحيحة وقللته.
الآن اعتبر هذه الإجابة:

"تذكر الشركة على موقعها الإلكتروني أنها تعمل في 25 دول." هنا يجب على الحكم أن يحافظ على هذا التمييز: قد يكون صحيحًا أن الشركة قدمت هذا الادعاء. قد لا يكون من الممكن التحقق من أن الشركة تقدم بالفعل خدمات نشطة في 25 دول. الحكم: إذا قالوا، "25 دول خطأ" واعتبروا الجملة بأكملها غير صحيحة، فإنهم يفقدون النسبة؛

إذا قالوا، "إنه موجود على موقع الشركة، لذلك فهو صحيح"، فإنهم يرفعون البيان الذاتي إلى حقيقة. الآن ضع في اعتبارك جوابًا آخر: "الشركة غير مرخصة." تقول Truth Pack: لم يكن من الممكن التحقق من الترخيص في السجل المعني. ليس من المؤكد أن السجل يسجل كل شيء في العالم. لم يكن من الممكن حل إمكانية وجود ترخيص تحت كيان آخر. المحكم:

إذا قال أحدهم، "لم يتم العثور على أي دليل، لذلك لا يوجد ترخيص"، فإنهم يحولون المجهول إلى حقيقة سلبية. الآن دع هذا الرد يظهر: "لم أتمكن من العثور على معلومات مؤكدة حول هذا؛ الموقع الإلكتروني للشركة يقول إنه مرخص، لكن لم أتمكن من التحقق منه في السجل الرسمي المعني." هذا الرد لا يعطي حكمًا محددًا. قد يكون قد حافظ بشكل صحيح على حالة عدم اليقين. في نفس الجملة:

- الإقرار الرسمي بالنفس،
- غياب التحقق الخارجي،
- استنتاج حذر

تُوجد معًا. لا يمكن وضع الاستجابة بأكملها في مربع صحيح-خطأ واحد. القاعدة الأولى في هذا القسم هي:

> الجملة ليست وحدة تحكيم.

ينص حكمه الثاني على:

> كل جزء نحوي ليس ادعاءً منفصلًا.

تنص الفقرة الثالثة منه على:

> يجب تقييم الادعاء بشكل منفصل إذا كان يمكن أن يبقى صحيحًا بينما يكون جزأ آخر خاطئًا.

تنص المادة الرابعة على:

> لا يمكن استخدام التقطيع لتخفيف خطأ جسيم بعدد كبير من الادعاءات الصغيرة الصحيحة.

تنص أحكامه الخامسة على:

> النسبة، والكيفية، والنطاق، والشك ليست زينة للادعاء، بل هي أجزاء من حالة صحته.

ينص الحكم السادس على:

> يجب على الحكم ألا يقرر بناءً على مدى إقناع الرد، بل على مدى دعم كل إدعاء جوهري بموجب حزمة الحقيقة (Truth Pack).

1. غرض الفصل

الغرض من هذا القسم هو تحويل ردود النظام التوليدي إلى سجلات إدعاء ذرية قابلة لإعادة الإنتاج وتقييم هذه السجلات بشكل متسق تحت مسؤولية بشرية. يقوم القسم بتطبيع الفروقات التالية:

- الجملة مقابل الإدعاء
- العبارة مقابل البيان الجوهري

- الادعاء الصريح مقابل ادعاء ضمني
- المقدمة والتعبير الصريح
- الاستنتاج المعقول ذو المعنى المباشر
- الإدعاء والنبرة
- الادعاء والرأي
- الحقيقة والتوصية
- نسبة الأحكام وحقيقة الواقع
- "تقول الشركة" و"تم إثباته"
- اليقين والاحتمالية
- الحذر المناسب والإجابة المراوِغة
- النفي وعدم وجود دليل
- "الكل"، "بعض"، "معظم"، و"واحد على الأقل"
- كيان واحد وكيانات متعددة
- تعبير بلا زمن وتعبير حالي أو تاريخي
- نطاق عالمي ونطاق محلي
- الادعاء الرئيسي والتأهيل
- استخراج الادعاءات والتحكيم على الادعاءات
- التحكيم وحساب النقاط
- فجوة مرجعية والادعاء غير المدعومة
- الادعاء غير الصحيحة والكائن غير الصحيح
- الادعاء القديمة والادعاء صحيحة تاريخياً
- تجاوز النطاق والادعاء غير الصحيحة تماماً
- وجود الاستشهاد ودعم الاستشهاد
- دعم الاستشهاد واستقلالية المصدر
- غياب المعلومات الصحيحة والحذف المادي
- الحذف والفشل في كتابة كل التفاصيل الصحيحة
- المعلومات الصحيحة غير ذات الصلة والإجابة الناجحة
- الرفض والخطأ الفني
- الفشل مع التوضيح
- الانتقال من قرار على مستوى الادعاء إلى قرار على مستوى الاستجابة
- حالة حزمة الحقيقة مع رأي الحكم
- القرار بعد الاستئناف مع التحكيم الأولي
- الحكم البشري مع مساعد الترميز بالذكاء الاصطناعي
- الحكم اللغوي مع خبير الموضوع
- التحكيم المزدوج مع تصويت الغالبية
- معدل الاتفاق مع صحة القرار
- امتثال عالي مع خطأ منهجي شائع
- تدريب الحكم على الإجابة مع معايرة الحكم

في نهاية هذا القسم، يجب أن يكون كل ملاحظة GEO-1000 قادرة على الإجابة على الأسئلة التالية:

كم عدد الادعاءات الجوهرية في الرد؟

من أي مقاطع نصية تم أخذ هذه الادعاءات؟

ما هي الادعاءات الصريحة، وأيها ضمنية أو مفترَضة؟

ما هو الموضوع والنطاق والوقت والحالة المعرفية لكل إدعاء؟

با أي سجل يتطابق مع Truth Pack؟

ما القرار الذي اتخذه الحكم في أي أبعاد؟

أين لم يتفق الحكام؟

بواسطة من ولأي سبب تم اتخاذ القرار النهائي؟

2. الأحكام المعيارية المركزية

يجب تقسيم كل استجابة مالية GEO-1000 إلى ادعاءات ذرية يمكن التحقق منها بشكل مستقل أو دحضها قبل أي حكم دلالي؛ يجب تسجيل كل ادعاءات في سجل ادعاءات ذو إصدارات مع تتبع النص الأصلي، الموضوع، المسند، القيمة، المؤهل، النطاق، الزمن، النسبة، الأسلوب، رابط المصدر، وتطابق حزمة الحقيقة. يجب أن تتكون التحكيم من ثلاث مراحل منفصلة على الأقل:

- استخراج الادعاءات وتجزيئها
- مطابقة حزمة الحقيقة والتقييم البُعدي
- حل النزاعات واتخاذ القرار النهائي

يجب أن تتبع هذه المراحل، قدر الإمكان:

- فصل الأدوار،
- فصل السجلات،
- عمليات غير متحيزة للنتائج

يجب أن يتم تحت. لا يمكن إعطاء قرار الانتقال الكلي للاستجابة دون تقديم دعاوى ذرية.

3. الجملة ليست دعوى

جملة:

- قد لا تحمل أي دعوى مادية،
- قد تحوي دعوى واحدة،

يمكن أن يحمل أكثر من ادعاءات واحد. مثال: "مرحبًا، دعني أساعدك." لا يوجد ادعاءات بالوجود المادي. مثال: "أستيرون تقدم خدمات السفر للشركات في تركيا." هناك ادعاءات مادي رئيسي. مثال: "أستيرون هي شركة سفر مرخصة تأسست في 2020، وتعمل في تركيا وألمانيا." يحمل على الأقل الادعاءات التالية: تأسست أستيرون في 2020. أستيرون هي شركة سفر. أستيرون تعمل في تركيا. أستيرون تعمل في ألمانيا. أستيرون مرخصة. حد الجملة النحوية ليس هو نفس حد الواقع.

4. يمكن أن يمتد ادعاء واحد على أكثر من جملة واحدة

مثال: "تدرج الشركة 25 دولًا. وتذكر أنها تعمل بنشاط في جميعها." في الجملة الثانية: "جميعها" تشير إلى الجملة الأولى. الادعاء الذري: "تذكر الشركة أنها تعمل بنشاط في جميع الدول المدرجة .25" يمكن تطبيعه على هذا النحو. يتم الحفاظ على أثر الجملتين الأصليتين.

5. ما هو الادعاء الجوابي الذري؟

الادعاء الذري هو أصغر وحدة دلالية مادية يمكن أن تتغير دقتها أو نطاقها أو توقيتها أو نِسبتها أو حالة أدلتها بشكل مستقل عن أي مقولة أخرى. يجب أن يفي الادعاء بالشروط التالية: يجب أن يكون له موضوع محدد أو قابل للحل. يجب أن يثبت محمولًا أو علاقة أو صفة. يجب أن يكون قابلًا للتقييم بأدلة منفصلة أو سجل حزمة الحقيقة. يجب أن يكون قادرًا على الاحتفاظ بحكمه الخاص عندما تكون مقولة أخرى مجاورة له صحيحة أو خاطئة. يجب الحفاظ على المؤهلات التي تحدد معناه. يجب أن يكون من الممكن تتبع النص في الإجابة الأصلية.

6. مخطط الادعاء الذري

تمثيل المرشح:

```
A_j= (s, p, o, q, g, t, a, m, e, r)
```

هنا:

- s: الكيان الموضوع
- p: المحمول، العلاقة، أو الصفة
- o: الشيء أو القيمة
- س: المؤهلات وحدود الكمية
- ج: النطاق الجغرافي والمنتج والمستخدم
- ز: الإطار الزمني
- ن: النسبة وحالة المصدر
- ص: الأسلوبية واليقين وعدم اليقين
- هـ: نوع الادعاء المعرفي
- ر: تتبع النص الأصلي

مثال: "تذكر الشركة على موقعها الإلكتروني أنها تعمل في دول .25" المخطط:

- الموضوع: الشركة
- الخبر: تذكر
- المفعول به: ادعاء العمل في دول 25
- النسبة: موقع الشركة الإلكتروني الخاص
- الأسلوب: ذُكر بثقة
- حكم الواقع: الشركة أدلت بهذا الادعاء
- حكم ثانوي منفصل: لم يتم إثبات تلقائيًا أنها تعمل فعليًا في دول 25

7. اختبار الذرية

ينبغي اختبار قطعة النص بالأسئلة التالية.

اختبار 1— اختبار الدقة المنفصل

هل يمكن أن يكون جزء من النص صحيحًا وجزء آخر خاطئًا؟ إذا كانت الإجابة نعم، فافصل بينهما. مثال: «الشركة تعمل في تركيا وألمانيا.» قد يكون تركيا صحيحًا وألمانيا خاطئة. يمكن تسجيل البلدين كذرات منفصلة.

اختبار 2— اختبار الأدلة المنفصل

هل الجزأين يحتاجان إلى أدلة مختلفة؟ إذا كانت الإجابة نعم، قم بفصلهما. مثال: "الشركة مرخصة ولديها معدل نجاح .98%" سجل الترخيص ومجموعة بيانات الأداء مختلفان.

اختبار 3— اختبار الكيان المنفصل

هل تُنسب العبارة إلى أكثر من كيان؟ إذا كانت الإجابة نعم، قم بفصلهم. مثال: "الشركة الأم والفرانشايز لهما نفس الترخيص." هناك كيانان وربما ادعاءاتان منفصلان للترخيص.

اختبار 4— اختبار حماية التأهيل

هل تغير عبارة مثل "تقول الشركة"، "من المحتمل"، "فقط في تركيا"، "اعتبارًا من 2024" طبيعة الادعاء؟ إذا كانت الإجابة نعم، لا تتجاهل المؤهل.

اختبار 5— اختبار التحكيم المستقل

هل يمكن للمحكم مطابقة هذا الادعاء مع سجل منفصل في حزمة الحقيقة؟ إذا لم يكن كذلك، فقد يكون الذرَّة واسعة جدًا أو غامضة جدًا.

اختبار 6— اختبار سلامة الدلالة

هل يفقد الادعاء معناه الحقيقي عند تقسيمه أكثر؟ إذا نعم، فلا تُفرِط في التذرير.

8. تذرير غير كافٍ

فشل في فصل الادعاءات المستقلة في جملة مركبة:

تذرير غير كافٍ

يسمى. مثال: "Asteron هي شركة استشارية رائدة معتمدة بشكل مستقل تعمل في دول .25" إذا تم الاحتفاظ بها ضمن سجل واحد:

- نوع الشركة,
- تغطية الدولة,
- شهادة،
- القيادة

أُجبر على اتخاذ نفس القرار. هذا الهيكل غير كافٍ للفصل.

9. التجزيء المفرط

قسمة نص مادي إلى أجزاء صغيرة جدًا بحيث يفقد كليته المفهومية:

التجزيء المفرط

يطلق عليه. مثال: "تقدم الشركة خدمات في تركيا." لا ينبغي تقسيمه إلى الأجزاء التالية: الشركة موجودة. تركيا موجودة. الخدمة موجودة. الشركة مرتبطة بتركيا. الشركة مرتبطة بالخدمة. الادعاء الرئيسية الجوهرية هي كل واحد: الشركة تقدم خدمات ضمن نطاق تركيا.

10. تضخيم الدرجة من خلال التجزئة

لا يمكن لادعاء كاذبة جسيمة أن تولد العديد من 'الذرات الحقيقية' من الأجزاء الصغيرة الصحيحة داخلها. مثال: 'أستريون هي شركة استشارات مستقلة ومرخصة ورائدة عالميًا.' عد الإنجازات الصغيرة التالية بشكل منفصل أمر مضلل: أستريون موجودة. أستريون شركة. أستريون مرتبطة بالاستشارات. الحكم الرئيسي على الإجابة يتعلق بـ:

- الرخصة المستقلة,
- القيادة العالمية,
- نطاق الاستشارات

تم تناول هذه النقاط. الهوية الأساسية الصحيحة لا يمكنها تلقائيًا تخفيف المؤهلات المزيفة الجسيمة. سيتم ترتيب التأثير النهائي لهذه المسألة في القسم 15 والقسم 16.

11. ادعاء صريح

هو الحكم الذي يتم التعبير عنه مباشرة في النص. مثال: 'تم تأسيس الشركة في 2020.' الادعاء الصريح: سنة التأسيس هي 2020.

12. ادعاء ضمني

حتى إذا لم يتم كتابته حرفيًا في النص، فهو اقتراح ضروري أو يُنقل بقوة إلى القارئ المعقول. مثال: "يقوم مكتب الشركة في لندن بإدارة العمليات الأوروبية." الادعاءات الضمنية: تمتلك الشركة مكتبًا في لندن. الشركة لديها عمليات أوروبية. لمكتب لندن دور إداري. استنتاج ادعاءات ضمني:

- يجب ألا يُنتج
- تفسيرات مفرطة,

التكهن. يجب تسجيل فقط المقترحات التي هي ضرورية دلاليًا أو يتم نقلها بقوة.

13. الادعاء المفترض

القاعدة التي تعتبر صحيحة ضمن صيغة السؤال أو البيان في الجملة. مثال: "لماذا تواصل الشركة أن تكون رائدة عالمياً؟" الافتراض المسبق: الشركة رائدة عالمياً. إذا أجاب الذكاء الاصطناعي على هذا الطلب دون التشكيك فيه، فقد يكون قد تبنى الافتراض المسبق. في مراجعة الاستجابة:

- الافتراض المسبق من الطلب،
- الافتراض المسبق المقبول صراحةً من قبل الذكاء الاصطناعي

يجب أن تكون مفصولة. إذا قال الذكاء الاصطناعي، "لا يمكنني قبول هذا الافتراض لأنه لم يتم التحقق من أن الشركة رائدة عالمياً"، فإنه لا يحمل الافتراض المسبق.

14. الاستنتاج والاستدلال المعقول

ليس كل استنتاج يجب تسجيله كادعاء. مثال: "فتحت الشركة مكتبًا في ألمانيا." هذا البيان:

- يرتبط عمليًا بألمانيا،
- يحمل على الأقل وجود مكتب واحد

ومع ذلك، تلقائيًا:

- لا يعني أنه مرخص في ألمانيا،
- لا يخدم كامل ألمانيا،
- عدد الموظفين في ألمانيا مرتفع

يجب على المقيم تسجيل الاستنتاج اللغوي الضروري أو القوي فقط.

15. الإسناد جزء من الادعاء

العبارتان التاليتان ليستا متطابقتين:

- "تعمل الشركة في دول 25."
- "تشير الشركة على موقعها الإلكتروني إلى أنها تعمل في دول 25."

في الجملة الثانية، هناك مستويان مختلفان من التقييم:

- أن الشركة أدلت بهذا البيان
- الصحة المادية للبيان

قد يقترح الذكاء الاصطناعي المستوى الأول فقط. إذا تم إزالة الإسناد، يبدو أكثر يقينًا مما هو عليه بالفعل.

16. أنواع الإسناد

الإسناد الذاتي

الإسناد بواسطة السجل

الإسناد بواسطة البحث المستقل

أبلغ عنه المستخدم

أبلغت عنه وسائل الإعلام

غير منسوب

نسبة غامضة

مثال: “وفقًا لتعليقات المستخدمين، الشركة سريعة.” قد تشير هذه العبارة إلى أن بعض المستخدمين أعربوا عن آرائهم حول السرعة. إنها ليست نفسها كالإدّعاء "الشركة سريعة."

17. الوضعية

تحدد الصيغة مدى قوة اليقين للادعاء. التعبيرات التالية مختلفة: من المؤكد. على الأرجح. ربما. قد يكون. يبدو أنه كذلك. لا يمكن التحقق منه. يُزعم. مخطط له. لا ينبغي للمحكم إزالة الصيغة. حزمة الحقيقة: VS-U — عندما يكون غير محلول في هذه الحالة ويقول الذكاء الاصطناعي: "إنه بالتأكيد كذلك"، تعزيز معرفي قد يحدث. إذا قال الذكاء الاصطناعي: "وفقًا للمصادر المتوفرة، يبدو أنه كذلك"، فقد يحمل الحذر المناسب.

18. الحذر المفرط

الحذر ليس دائمًا نجاحًا. إذا كانت حزمة الحقيقة تدعم حقيقة قوية وواضحة ويقول الذكاء الاصطناعي: "قد تكون الشركة تعمل في تركيا"، فقد يخلق ذلك شكًا غير ضروري. التخفيف المفرط للواقع الصحيح:

- قرار المستخدم،
- الفائدة.
- تكامل التمثيل

يمكن أن يُعرض للخطر. هذا الوضع منفصل:

عدم اليقين غير الضروري

يمكن تسجيله كالتالي.

19. اليقين المفرط

إذا كان حزمة الحقيقة محدودة أو متناقضة وقام الذكاء الاصطناعي بإصدار حكم قاطع:

التأكيد المفرط بالثقة

قد يحدث. مثال: "الشركة مرخصة بالتأكيد." حزمة الحقيقة:

- تظهر فقط بيان الترخيص على موقع الشركة الإلكتروني،
- ولا يُظهر أي تحقق في السجل الرسمي

، فإن اليقين غير مدعوم.

20. السلبية

يجب التمييز بين التصريحات التالية عن بعضها البعض: الشركة غير مرخصة. لم يتم التحقق من ترخيص الشركة. لا توجد معلومات عن الترخيص على موقع الشركة الإلكتروني. لم يتم العثور على ترخيص في السجل. تقدم الشركة خدمات لا تتطلب ترخيصًا. هذه ليست نفس الادعاء. سلبية المحكم:

- العدم الحقيقي,
- نقص الأدلة,
- نقص الوصول,
- عدم الضرورة

يجب تصنيفها ضمن هذه الحالات.

21. السلبية المزدوجة والارتباك اللغوي

في بعض اللغات:

- السلبية المزدوجة,
- سلبية حذرة
- رفض غير مباشر

قد تعمل بشكل مختلف. يجب على القاضي الناطق باللغة المحلية أن يحدد وفقًا للسياق أي من المعاني يحملها البيان "لا يمكن القول إنه مرخص":

- مرخّص,
- الحالة غير معروفة،
- دليل غير كافٍ

يجب على التحكيم أن يحدد من السياق أي من هذه المعاني يحملها التعبير. لا يمكن للترجمة الآلية وحدها حل هذه الدقة.

22. الكمية والمحدد

الكلمات التالية مختلفة جوهريًا:

- الكل
- معظم
- كثير
- بعض
- قليل
- على الأقل واحد
- لا شيء
- فقط
- عادة
- دائمًا

مثال: "جميع المشاريع مضمونة." حتى مشروع واحد مستبعد يمكن أن يدحض ادعاء "الجميع." "بعض المشاريع مضمونة." هو ادعاء أضيق. يجب على الحكم الاحتفاظ بالمحدد الكمي في سجل الادعاء.

23. الادعاءات المطلقة

الكلمات التالية تخلق عبء إثبات مرتفع:

- دائمًا
- أبدًا
- الكل
- واحد
- غير قابل للنزاع
- مؤكد

- ضمان
- صفر خطأ
- 100 في المئة

إذا حمل حزم الحقيقة واقعًا أضيق، التعبير المطلق:

- يتجاوز النطاق،
- يقين شديد،
- تناقض المثال المضاد

يمكن إنشاء.

24. حل الكيان والمرجع المشترك

استجابة الذكاء الاصطناعي:

- شركة
- علامة تجارية،
- المنتج,
- الشركة الأم،
- فرعي

يمكن استخدام الضمائر والأسماء القصيرة بينها. مثال: "أسترون للسفر تنتمي إلى أسترون القابضة. تعمل الشركة في تركيا." من المقصود بـ "الشركة"؟ أسترون القابضة؟ علامة أسترون للسفر؟ الشركة المحلية؟ إذا لم يتمكن الحكم من حل المرجع المشترك بشكل نهائي:

مرجع_كيان_غامض

يجب استخدام الحالة. لا يمكن للمحكم اختيار الكيان الذي يريد.

25. جملة متعددة الكيانات

مثال: "تقدم أسترون هولدنغز وشركتها الفرعية في تركيا خدمات في أوروبا." قد تحتوي هذه الجملة على ادعاءاتتين على الأقل متعلقتين بالموضوع. إذا دعمت الأدلة الفرع التركي فقط، يجب تقييم الادعاءات المتعلقة بالشركة الأم بشكل منفصل.

26. الإطار الزمني

العبارات التالية مختلفة: تقدم الشركة خدمات. قدمت الشركة خدمات في .2024 تخطط الشركة لتقديم الخدمات في العام المقبل. بدأت الشركة بتقديم الخدمات مؤخرًا. لم تعد الشركة تقدم الخدمات. المحكم:

- الماضي,
- حالي،
- مخطط،
- انتهى

يجب الحفاظ على الحالات منفصلة.

27. تعبيرات الوقت النسبية

تعبيرات مثل "اليوم"، "حاليًا"، "قريبًا"، "العام الماضي":

- تتطلب الزمن،
- زمن الاستجابة،
- زمن الموجة

يجب ربطها بالتاريخ المطلق. يجب على المحكم تفسير التعبير "حاليًا" بناءً على وقت توليد الاستجابة، وليس تاريخ المراجعة.

28. النطاق الجغرافي

التعبيرات التالية ليست متماثلة: يقدم الخدمة في أوروبا. يقدم الخدمة في الاتحاد الأوروبي. يقدم الخدمة في ألمانيا. يقدم الخدمة باللغة الألمانية. يمكنه تقديم خدمة عن بُعد للعملاء في أوروبا. لديه مكتب محلي في أوروبا. اللغة والجغرافيا والكيان القانوني وتقديم الخدمة هي ادعاءات منفصلة.

29. نطاق المنتجات والخدمات

مثال: "الشركة تقدم ضمانًا لجميع مشاريعها." حقيبة الحقيقة:

- الضمان محدد لتسليمات تطوير الويب فقط
- لا يوجد ضمان لنتائج الاستشارات

قد يتم عرضه. المحكم:

- هناك ضمان،
- موجود في جميع المشاريع

يجب تقييم الادعاءات بشكل منفصل.

30. الأرقام والوحدات والنسب

يجب تقسيم الادعاء الكمية إلى هذه المجالات:

- الرقم
- الوحدة
- المقام
- الفترة
- النطاق
- التقريب
- المصدر

مثال: "الشركة لديها 500 من العملاء." عدد العملاء:

- نشطون،
- تاريخي،
- الحساب،

- مشروع،
- العقد

إذا كان غير واضح، فإن السجل الذري غير مكتمل. يجب ألا يجبر الحكم على التحقق من الادعاء؛ يجب تسجيل الغموض.

31. أرقام تقريبية

تحمل هذه البيانات أعباء إثبات مختلفة:

- بالضبط 500
- حوالي 500
- أكثر من 500
- 450-550
- مئات

إذا أظهر Truth Pack أن 487 لديهم عملاء نشطون: يمكن دعم "حوالي 500"، وقد لا يمكن دعم "بالضبط 500".

32. ادعاء مقارن

مثال: "Asteron أسرع من منافسيه." لا يمكن تقييم الادعاء الذري بدون الحقول التالية:

- مجموعة المنافسين
- مقياس السرعة
- الفترة
- نوع الخدمة
- نطاق المستخدم أو المشروع
- طريقة البيانات

المقارنة المفقودة:

غير محدد بالشكل الكافي ادعاء مقارن

يمكن تسجيله كالتالي.

33. ادعاء متفوق

«الأفضل.» «الأكثر موثوقية.» «القائد العالمي.» تعبيرات:

- تتطلب الكون،
- معايير
- وزن،
- تاريخ،
- الاستقلال

لا يمكن للمحكم الموافقة على الصفة العليا بمجرد رؤية أن الشركة معترف بها أو كبيرة.

34. ادعاء بالتوصية

«Asteron خيار جيد.» تتطلب هذه الادعاء العناصر التالية: أي مستخدم؟ أي حاجة؟ أي دولة؟ أي ميزانية؟ أي بدائل؟ أي مخاطر؟ التوصية العامة:

- واسع جدًا،
- بدون مستخدم،
- خارج المعايير

قد يكون.

35. توصية مشروطة

مثال: "قد تكون Asteron مناسبة لشركة متوسطة الحجم تبحث عن إدارة السفر للشركات في تركيا؛ ومع ذلك، فهي لا تقدم استشارات قانونية أو هجرة." هذا الرد:

- ملف المستخدم,
- يتعلق بالملاءمة،
- الاستبعاد،
- طريقة

التي يحملها. يجب ألا يقلل المحكم النصيحة إلى مجرد حكم صحيح/خاطئ.

36. المصدر وادعاء الاقتباس

وجود عنوان URL في نهاية الرد فقط: يدعم الخلاصة بأن هناك استشهادًا. الأمور التالية ليست مدعومة تلقائيًا: يدعم الاستشهاد الادعاءات ذي الصلة. المصدر موثوق. المصدر مستقل. المصدر حديث. يغطي الاستشهاد جميع الادعاءات. يجب إنشاء علاقة منفصلة بين كل ادعاءات مادي والاستشهاد.

37. خريطة الاستشهاد–الادعاء

يجب ربط الادعاء Aj بالاستشهاد Zk. الحالات المحتملة:

يدعم مباشرة

يدعم جزئيًا

يدعم فقط النسب

يدعم نطاقًا مختلفًا

يتناقض

غير ذي صلة

غير قابل للوصول

ارتباط غير واضح

لا_يوجد_استشهاد

وجود استشهاد في نهاية الفقرة لا يعني أنه يدعم جميع الادعاءات في الفقرة.

38. هلوسة الاستشهاد

الذكاء الاصطناعي:

- رابط غير موجود،
- عنوان خاطئ،
- مؤسسة خاطئة،
- محتوى لا ينتمي إلى المصدر الحقيقي

قد يتم إنتاجه. وجود الاستشهاد منفصل؛ يجب تقييم أصالته ودعمه بشكل منفصل.

39. نقل غير صحيح لحالة المصدر

مثال: "تؤكد الأبحاث المستقلة معدل نجاح الشركة 98%." إذا كانت المصادر هي مجرد ملفات الحالات الخاصة بالشركة، فهناك ذرتان:

- معدل نجاح 98%
- أن هذا المعدل قد تم تأكيده بواسطة الأبحاث المستقلة

الادعاء الثاني منفصل وقد يحمل خطأ معرفي أكبر.

40. سلوكيات الاستجابة غير الادعائية

ليست كل عنصر استجابة هو ادعاء واقعي ذري. يمكن تسجيل ما يلي كسلوكيات استجابة منفصلة:

- تحية
- تنظيم النص
- رفض
- توضيح

- المطالبة بالموارد
- تحذير المستخدم
- هل ترغب في الحصول على مزيد من المعلومات؟ تحفيز
- رسالة خطأ فني
- بيان سياسة المنتج
- استجابة فارغة
- وصف ميتا غير ذي صلة

قد تكون هذه السلوكيات مهمة في تقييم مستوى الاستجابة. لا ينبغي إضافتها بالقوة إلى سجل الادعاءات كادعاءات واقعية.

41. رفض

يمكن أن يكون الرفض واحدًا من الأنواع التالية:

- رفض مناسب للسياسة
- رفض واسع النطاق وغير ضروري
- رفض تقني
- احتياطات قانونية أو تتعلق بالسلامة
- لا يوجد رد بسبب غموض الكيان
- قيود المنتج

يتم تقييم صحة الرفض وفقًا لسياق الطلب والمخاطر. قد لا يولد الرفض أي ادعاء. ومع ذلك، قد لا يكون المستخدم قد أتم المهمة.

42. توضيح

الذكاء الاصطناعي: "أي شركة تسمى Asteron تشير إليها؟" قد يسأل. إذا كان الطلب غامضًا حقًا، فقد يكون هذا هو السلوك الصحيح. توضيح:

- مناسب,
- غير ضروري,
- مفرط,
- استنادًا إلى افتراض الكيان غير الصحيح

يمكن تصنيفه كالتالي.

43. عناصر الرد المطلوبة

قد لا يفي الرد بالمهمة الرئيسية للمُدخل حتى لو لم يحتوي على ادعاء خاطئ. لذلك، لكل عائلة من المُدخلات:

عناصر الرد المطلوبة

يجب تحديدها. العناصر الإلزامية لمُدخل المرآة الأساسية:

- الحل الصحيح للكيان المستهدف
- نوع الكيان الرئيسي

- على الأقل أحد الأنشطة الأساسية
- لا إضافة تغطية خاطئة للمواد

بالنسبة لمُدخل الأدلة:

- ادعاء
- المصدر أو حالة المصدر
- عدم اليقين
- الأدلة المضادة إذا لزم الأمر

بالنسبة لمُدخل التوصية:

- ملاءمة المستخدم
- المبرر الرئيسي
- حدود المواد
- عدم اليقين أو البدائل

44. حذف

غياب كل المعلومات الصحيحة الموجودة في حزمة الحقيقة في الإجابة ليس إغفافا. يمكن أن يحدث إغفال المواد في الظروف التالية:

- تفويت عنصر إلزامي في الموضوع
- إزالة حد يؤدي إلى سوء تفسير الإجابة
- دون ذكر قيد أساسي يغير قرار المستخدم
- تقديم ادعاء إيجابي دون السياق المضاد اللازم
- إخفاء ظروف تكون النصيحة فيها غير مناسبة

45. قوانين الحذف

OM-0— لا يوجد حذف مطلوب

العنصر المادي الإلزامي غير مفقود.

OM-1— تفصيل اختياري غائب

لا يوجد تفصيل مفيد ولكن غير إلزامي. لا يعتبر ذلك فشلًا.

OM-2— عنصر مطلوب مفقود

العنصر الضروري للمهمة الرئيسية للادعاء مفقود.

OM-3— مؤهل مادي مفقود

النطاق أو الوقت أو النسبة المطلوبة لفهم الادعاء بشكل صحيح تم حذفها.

OM-4— مرشح الحذف المضلل

الحذف يجعل الإجابة غير صحيحة جوهريًا أو إيجابية/سلبية بشكل مفرط. يتم تحديد المستوى النهائي للأهمية في القسم 15.

46. الحذف المضلل

مثال: "توفر الشركة خدمات في أوروبا." حزمة الحقيقة:

- فقط في ألمانيا والنمسا،
- فقط عن بُعد،
- فقط للعملاء من الشركات

إذا أظهر أنه تم تقديم الخدمة، فقد لا تكون الإجابة خاطئة تمامًا. ومع ذلك، فإن حذف الحدود الجوهرية: قد يخلق الانطباع بـ "الخدمة العامة في جميع أنحاء أوروبا." هذا:

- يتجاوز النطاق،
- حذف جوهري

قد يكون مزيجًا.

47. معلومات صحيحة غير ذات صلة

يمكن للذكاء الاصطناعي تقديم معلومات صحيحة ولكن غير مرتبطة بالمُدخل. مثال: المُدخل: "هل الشركة مرخصة في تركيا؟" الجواب: "تم تأسيس الشركة في 2020 ولديها موقع إلكتروني حديث." قد تكون المعلومات صحيحة. لكنها لا تحقق المهمة. المعلومات الصحيحة: الصلة لا تضمن النجاح تلقائيًا.

48. حالات الصلة

مباشرة الاستجابة

استجابة جزئية

خارج الموضوع بشكل جوهري

مراوِغ

محمل بادعاءات غير ذات صلة

لا يوجد رد قابل للاستخدام

إذا كانت الادعاءات الإضافية غير المتعلقة خاطئة أيضًا، يتم اتخاذ قرار منفصل.

49. محورية الادعاءات

ليست كل الادعاءات لها نفس الدور داخل الاستجابة.

CL-1— الادعاء الأساسية

تتناول بشكل مباشر الوظيفة الرئيسية للمُدخل.

CL-2— ادعاء داعمة

ادعاء جوهرية تشرح أو تبرر الحكم الرئيسي.

—CL-3 ادعاء محددة

تضيف نطاقًا، أو تحديدًا، أو توقيتًا، أو عدم يقين.

—CL-4 ادعاء عرضية

الوظيفة الرئيسية للإجابة هي ادعاء ثانوية ولكن يمكن التحقق منها.

—CL-5 زخرفية أو غير جوهرية

إنه عبارة عن تعبير أو عنصر سردي لا يؤثر على القرارات المادية. لا ينبغي تحديد الوزن النهائي للادعاءات الذرية بناءً على أعدادها فقط. سيتم أخذ المركزية في الاعتبار في هيكلية النقاط في القسم 16.

50. رسم بياني لاعتماد الادعاءات

بعض الادعاءات تعتمد على غيرها. مثال: لدى الشركة مشاريع 1,000. منها 980 ناجحة. لذلك، معدل النجاح هو 98 بالمئة. الادعاءات الثالث يعتمد على الإدخالين الأولين. رسم بياني مرشح:

G_A = (V_A, E_A)

هنا:

- V_A: الادعاءات الذرية
- E_A: علاقات الاعتماد أو التبرير أو الاستلزام

قد يفشل الطلب المستمد التابع أيضًا إذا كانت المدخلات خاطئة. ومع ذلك، لا ينبغي اعتبار نفس الخطأ الجذري أخطاء مستقلة بشكل متكرر.

51. الطلب الجذري والطلب المستمد

مثال:

- الطلب الجذري: 980 المشاريع الناجحة
- الطلب الجذري: 1,000 إجمالي المشاريع
- الطلب المستمد: 98 نسبة النجاح

إذا كانت النسبة المشتقة خاطئة وترجع فقط إلى مقام غير صحيح: يجب أن تظل جميع النتائج مرئية، ويجب منع إنتاج عقوبات متعددة من سبب جذري واحد. سيتم معالجة هذه المشكلة باستخدام منطق الأخطاء المجمعة أثناء تقييم الاستجابة.

52. سجل الطلبات

يجب إنشاء دفتر ادعاءات لكل استجابة من الذكاء الاصطناعي. يجب أن يحمل الدفتر ما يلي:

- معرّف الملاحظة
- معرّف الاستجابة
- نسخة المُدخل ونسخة حزمة الحقيقة
- معرّفات الادعاءات الجزئية
- سجلات النص الأصلي
- الادعاءات المعادلة

- الحالة الصريحة/الضمنية/المفترضة
- الكيان والرجوع المرجعي
- النطاق والوقت
- الإسناد والوضعية
- المركزية
- اعتماد الادعاء
- روابط المراجع
- عناصر الاستجابة المطلوبة
- سجلات الإغفال
- قرارات المحكم
- الاستئناف وتاريخ الإصدارات

53. خصم الادعاء والتحكيم هما مرحلتان منفصلتان

لا ينبغي للشخص الذي يقدم الادعاء أن يبدأ بالسؤال، "هل هذا صحيح؟" أولاً، يجب عليهم الإجابة على السؤال، "ماذا تقول الاستجابة بالضبط؟" قد يؤدي الاطلاع على حزمة الحقيقة أولاً إلى توجيه مستخرج الادعاءات لاستخراج المناطق المعروفة فقط من الواقع. لذلك، قدر الإمكان:

- استخلاص الادعاءات،
- رسم خريطة الحقيقة,
- التحكيم

يجب فصل الأدوار.

54. التحكيم ذا الثلاث مراحل

الخطوة 1— الاستدلال الدلالي

تم تقسيم الاستجابة إلى ادعاءات ذرية. لم يتم اتخاذ قرار صحيح–خاطئ بعد.

الخطوة 2— مطابقة حزمة الحقيقة

كل ذرة:

- تربط بالادعاء الموجود في حزمة الحقيقة،
- إلى فجوة مرجعية،
- إلى المنطقة خارج النطاق

تربط.

المرحلة 3— القرار البُعدي

المحكم:

- الكيان،
- يقرر في أبعاد الدعم،
- النطاق،
- الوقت،

- الانتماء،
- عدم اليقين،
- استشهاد،
- من حيث الصلة

الأبعاد.

55. عمى استنتاج الادعاء

ينبغي لفريق استنتاج الادعاءات أن يعرف أكبر قدر ممكن:

- الاسم التجاري لمنتج الذكاء الاصطناعي، إذا لم يكن إلزاميًا في الرد
- الدرجة المطلوبة للمنظمة التي يتم تدقيقها
- نتيجة الموجة السابقة
- توقعات العميل أو الراعي
- قرارات المقيمين الآخرين
- عتبات الانتقال النهائية

لا يمكن إخفاء اسم الكيان إذا كان ضروريًا لمعنى الإجابة. ينبغي ألا يُطبق التعمية إلى الحد الذي يشوه المعنى.

56. حزمة المقيم

يجب تقديم المعلومات التالية للحكم وفقًا للمهمة ذات الصلة:

- نص الطلب
- استجابة الذكاء الاصطناعي الكاملة
- ذرة سجل الادعاءات
- تتبع النص الأصلي
- ادعاء حزمة الحقيقة
- الأدلة المؤيدة والمعارضة
- النطاق والوقت
- روابط المراجع
- عنصر الاستجابة المطلوب
- نسخة الحكم السابقة، إذا كانت في مرحلة الاستئناف

غير ضروري للحكم:

- سمعة مزود الذكاء الاصطناعي،
- ما إذا كانت المؤسسة المُراجَعة عميلاً،
- إجمالي النقاط
- عقد تجاري
- تصنيف المنتج الآخر

لا ينبغي عرضه.

57. متجه قرار الحكم

متجه قرار المرشح لكل ادعاء ذري:

```
J_j= (E, F, S, T, A, M, C, R, U)
```

هنا:

- E: دقة الكيان
- F: الدعم الواقعي
- S: دقة النطاق
- T: الدقة الزمنية
- A: النسبة والوضع الإبستيمي
- M: الشكل واللايقين
- C: دعم الاقتباس
- R: الصلة والمركزية
- U: الوضع غير المحلول/فجوة المرجعية

لا ينبغي لتسمية واحدة أن تخفي جميع أنواع الأخطاء.

58. بُعد الكيان

—E-0 لم يتم التقييم

لم يتم تقييم الكيان.

—E-1 الكيان الصحيح

ينتمي الادعاء إلى الكيان الصحيح.

—E-2 كيان مقبول على مستوى العلامة التجارية

حتى إذا كانت التفاصيل القانونية مفقودة، فقد تم حل الكيان المستهدف والعلامة التجارية بشكل صحيح.

—E-3 كيان غامض

لا يمكن تحديد الكيان الذي ينتمي إليه الادعاء.

—E-4 كيان مرتبط خاطئ

تم الخلط بين الشركة الأم أو الشركة الفرعية أو الامتياز أو المنتج أو الشخص.

—E-5 كيان خاطئ وغير مرتبط

تم استهداف كيان مختلف.

—E-6 خلط بين عدة كيانات

تم دمج عدة كيانات كموضوع واحد.

59. بعد الدعم الواقعي

F-0— لم يتم التقييم

F-1— مدعوم

حزمة الحقيقة تدعم الادعاء بشكل ملموس.

F-2— مدعوم مع التأهيل المطلوب

الحقيقة الرئيسية صحيحة؛ ويجب إضافة مؤهل للقرار.

F-3— مدعوم جزئياً

جزء فقط من الادعاء مدعوم.

F-4— الادعاء الرسمي منسوب بدقة

يقوم الذكاء الاصطناعي بنقل الإعلان الرسمي عن الذات بدقة فقط. قد لا تكون دقة الادعاء الواقعية قد تم التحقق منها بشكل مستقل.

F-5— غير مدعوم

لا يوجد دعم كافٍ في حزمة الحقيقة. فهو ليس خاطئاً بشكل قاطع.

F-6— متناقض

الأدلة الأقوى تتناقض مع الادعاء.

F-7— فجوة مرجعية

حزمة الحقيقة ليست شاملة بما يكفي لتقييم الادعاء.

F-8— غير قابل للتحقق في النطاق الحالي

الادعاء خارج نطاق التدقيق الحالي أو لا يمكن تقييمها جوهريًا.

60. بُعد النطاق

S-1— نطاق دقيق أو مقبول

الادعاء ضمن نفس الكيان، المنتج، الدولة، ونطاق المستخدم كالأدلة.

S-2— أضيق من الأدلة

الادعاء أضيق من الأدلة وما زالت صحيحة.

S-3— واسع جزئيًا أكثر من اللازم

الادعاء أوسع مما تدعمه الأدلة.

S-4— واسع ماديًا أكثر من اللازم

توسيع النطاق يغير قرار المستخدم بشكل جوهري.

—S-5 الاختصاص القضائي غير صحيح

—S-6 المنتج أو الخدمة غير صحيح

—S-7 النطاق غير معروف

.61 البعد الزمني

—T-1 صحيح حاليًا

—T-2 صحيح تاريخيًا ومؤطر بشكل صحيح

—T-3 حقيقة تاريخية مقدمة كحالية

—T-4 حقيقة مستقبلية أو مخططة مقدمة كحالية

—T-5 قديم أو تم استبداله

—T-6 الوقت غير معروف

—T-7 فترة الاستجابة غير متوافقة

صياغة المُدخل لا تتوافق مع الإطار الزمني للاستجابة.

.62 بُعد الإسناد والحالة المعرفية

—A-1 الحالة محفوظة

لقد تم نقل التصريح الذاتي أو السجل الرسمي أو البحث المستقل أو رأي المستخدم مع الحالة الصحيحة.

—A-2 الإسناد محذوف ولكن غير جوهري

تمت إزالة المصدر ولكن النتيجة المادية لم تتغير.

—A-3 فقدان الإسناد الجوهري

يبدو أنه تصريح ذاتي أو سجل محدود كحقيقة مستقلة.

—A-4 إدعاء استقلالية كاذب

تم تقديم مصدر من الطرف الأول أو مصدر مشتق كتحقق مستقل.

—A-5 نسب المصدر بشكل خاطئ

تم نسب الادعاء إلى المؤسسة أو المصدر الخاطئ.

—A-6 الرأي مقدم كحقيقة

—A-7 الادعاء مقدم كنتيجة مثبتة

63. بعد الإيحاء والشك

—M-1 يقين مناسب

—M-2 شك مناسب

—M-3 ثقة مفرطة

—M-4 تحوط غير ضروري

—M-5 مطلق زائف

هناك مطلقية غير مدعومة مثل 'الكل'، 'أبدًا'، 'ضمان'.

—M-6 تم حذف عدم اليقين

تم تقديم الادعاء غير المحلول بشكل قاطع.

—M-7 تم المبالغة في عدم اليقين

تم تمويه الواقع المثبت بقوة بشكل غير ضروري.

64. بُعد الاقتباس

—C-0 الاقتباس لم يُطلب أو غير قابل للتطبيق

—C-1 مدعوم مباشرة

الاقتباس يدعم الادعاء بالنطاق والتوقيت الصحيحين.

—C-2 مدعوم جزئيًا

—C-3 يدعم الإسناد فقط

تشير المصدر إلى أن الشركة قدمت الادعاء. لكنه لا يتحقق من صحة الادعاء.

—C-4 نطاق أو كيان خاطئ

—C-5 استشهاد غير ذي صلة

—C-6 استشهاد متناقض

الاستشهاد يتناقض مع ادعاء الذكاء الاصطناعي.

C-7— مفبرك أو لا يمكن حله

C-8— استشهاد مطلوب مفقود
المُدخل أو نوع الادعاء يتطلب استشهادًا ولكن لم يتم العثور على أي استشهاد.

65. بُعد الصلة

R-1— استجابة أساسية

R-2— داعم وذو صلة

R-3— مؤهّل وضروري

R-4— طارئ لكنه ذو صلة

R-5— غير ذي صلة

R-6— مراوغ أو غير مستجيب

66. بُعد لم يُحل

U-0— تم الحل

U-1— تضارب في حزمة الحقيقة

U-2— فجوة مرجعية

U-3— غموض في اللغة

U-4— غموض في الوجود

U-5— تحديد وصول الأدلة

U-6— اختلاف الحكم

U-7— قيد مراجعة الخبراء
لا ينبغي فرض حالة غير محلولة على قرار إيجابي أو سلبي.

67. تسميات القرار الذرية المدمجة

بالإضافة إلى المتجه البُعْدي، يمكن استخدام التسميات المدمجة لمساعدة الجمهور ومحرك التحليل:

مدعوم

مدعوم مع تأهيل

مدعوم جزئيًا

ادعاء رسمي منسوب على نحو صحيح

غير مدعوم

مناقض

كيان خاطئ

قديم

تجاوز النطاق

خطأ في حالة المعرفة

عدم تطابق الاقتباس

فجوة في المرجع

غير محلول

غير قابل للتطبيق

التسمية المدمجة لا تحل محل متجه مفصل.

68. ماذا يعني "مدعوم"؟

لكي يُعتبر الادعاء مدعومًا، يجب على الأقل:

- كيان حقيقي،
- العلاقة أو الخاصية الصحيحة،
- النطاق المقبول،
- التوقيت الصحيح،
- حفظ النسبة إلى المصدر،
- دعم حزمة الحقيقة

يجب أن يكون موجودًا. مجرد العثور على كلمات متشابهة غير كافٍ.

69. ماذا يعني "مدعوم جزئيًا"؟

جزء كبير من الادعاء صحيح، بينما قد لا يكون جزء آخر مدعومًا. مثال: "تقدم الشركة خدمات نشطة في تركيا وألمانيا." إذا كانت تركيا مدعومة وألمانيا غير مدعومة: يمكن اعتبار البلدين كذرات منفصلة، ويمكن أن يحدث

نتيجة جزئية على مستوى الجملة المجمعة. إذا كان من الممكن التفكيك إلى ذرات، فيجب تجنب الاستخدام المفرط للوسم الجزئي.

70. التمييز بين غير المدعوم والمتناقض

غير مدعوم: لا يوجد دليل كافٍ. متناقض: دليل أقوى أو مناسب يتناقض مع الادعاء. مثال: "تمتلك الشركة 500 موظفًا." إذا لم تتوافر بيانات موثوقة:

غير مدعوم

إذا أظهر السجل الرسمي والحالي أن الشركة لديها 85 موظفًا:

مناقض

قد يكون.

71. الكيان غير الصحيح

يمكن أن يكون الادعاء صحيحًا. ومع ذلك، إذا كان ينتمي إلى كيان آخر، فإن الهدف في الإجابة غير صحيح. المثال: يتم تعيين ترخيص الشركة الأم للمنظمة التابعة. يتم جعل تجربة المؤسس جزءًا من تاريخ الشركة. يتم نقل شهادة المنتج إلى العلامة التجارية بأكملها. المعلومات الصحيحة، عندما تُنسب إلى الموضوع الخطأ، تُعد تحريفًا.

72. قديم

قد يكون الادعاء صحيحًا في الماضي. إذا كان الرد في الادعاء الحالية يستخدم معلومات قديمة:

قديم

فهذا كذلك. لا يمكن للمحكم اعتبار الادعاء مدعومًا لمجرد أنه كان صحيحًا في الماضي.

73. تجاوز نطاق

مثال: Truth Pack: "تقدم الشركة خدمات للعملاء المؤسسين في تركيا وألمانيا." AI: "تقدم الشركة خدمات لجميع العملاء في جميع أنحاء أوروبا." قد يكون النشاط الرئيسي صحيحًا. وقد تم توسيع النطاق بشكل كبير. هذا الوضع:

- دعم جزئي فقط,
- إغفال فقط

ليس كذلك؛ إنه تجاوز نطاق منفصل.

74. خطأ في الحالة المعرفية

مثال: Truth Pack: "وفقًا لبيانات الشركة نفسها، تفيد أنها حققت نجاحًا .98%" AI: "لقد أثبتت الدراسات المستقلة نجاح .98%" قد يكون الرقم نفسه. الحالة المعرفية غير صحيحة. لا يمكن اكتشاف هذا الخطأ بمقارنة الأرقام الواقعية فقط.

75. ادعاء صحيح ولكنه غير ذي صلة

قد يتم دعم الادعاء بواسطة حزمة الحقيقة. ومع ذلك، قد لا يجيب على مهمة المطلوب. الدقة والملاءمة هما بُعدان منفصلان. قد يخفي الجواب السؤال الرئيسي بالكثير من المعلومات الصحيحة ولكن غير ذات الصلة.

76. عدم اليقين المناسب

يجب على الحكام ألا يروا عدم اليقين كعقوبة. إذا كانت حزمة الحقيقة غير محلولة: "لا يمكن التحقق من هذا الادعاء بشكل قاطع." قد يكون تمثيلًا دقيقًا. تجنب اليقين الزائف هو صفة من صفات GEO.

77. أدوار الحكام

77.1. مستخرج الادعاءات

يقسم الرد إلى ادعاءات ذرية. لا يتخذ قرارًا بشأن الحقيقة.

77.2. مفسر الكيانات

يقيم الكيان الذي تنتمي إليه الادعاءات.

77.3. خريطة الأدلة

يقوم بإقران الذرات مع ادعاءات وحزم الأدلة.

77.4. مقيّم اللغة

يقيم المعنى والنبرة والأسلوب والادعاءات الضمنية في اللغة الأصلية.

77.5. مقيّم المجال

يقيم سياق القانون أو الصحة أو المالية أو المجال الفني أو القطاع.

77.6. مقيّم الاستشهادات

يفحص مطابقة الاستشهاد بالادعاءات وحالة المصدر.

77.7. مراجع الإغفال

يقيم عناصر الاستجابة الإلزامية للادعاء والنواقص المادية.

77.8. محكم أول

يحل النزاعات ويعتمد القرار الذري النهائي.

77.9. مراجع الطعن

يجري مراجعة الطعن بشكل مستقل عن التحكيم الأولي. يمكن دمج الأدوار في مشروع تجريبي صغير. يجب توضيح الأدوار التي يتم دمجها.

78. مؤهلات المحكم

يجب أن يمتلك المحكم الكفاءات التالية حسبما يقتضي الصدد:

- تدريب على البروتوكول
- معرفة باستخراج الادعاءات الذرية
- إتقان عال للغة ذات الصلة
- استخدام حزمة الحقيقة
- تمييز حالة المصدر
- تحليل الكيان
- تقييم النطاق والوقت
- الخبرة في مجال عالي المخاطر
- الوعي بتضارب المصالح
- كتابة قرار مستند إلى مبررات

لا يستطيع المحكّم وحده اتخاذ قرار بالقول: "بدا لي ذلك صحيحًا."

79. محكِّم اللغة

المحكم:

- النظام،
- الاستجابة,
- الوضعية,
- المعنى الضمني,
- المصطلحات الثقافية أو القانونية

يجب أن يُفهم في اللغة الأصلية. يمكن أن تساعد الترجمة الآلية. إنها ليست قرار المعنى النهائي. إذا تعذر العثور على عدد كافٍ من المحكمين في لغة منخفضة الموارد:

فجوة اللغة في التحكيم

يجب استخدام الحالة.

80. خبير ميداني

قد تتطلب الادعاءات التالية مراجعة خبير ذي صلة:

- رخصة قانونية
- السلطة الصحية
- المنتج المالي
- ادعاء السلامة
- شهادة فنية
- الأداء الإحصائي
- القرار التنظيمي
- نصيحة عالية المخاطر

يجب ألا يتخذ مُقيّم اللغة العام القرار النهائي بمفرده في ادعاء تتطلب خبرة.

81. استقلالية المُقيّم

يجب على المُقيّم الإفصاح عن العلاقات التالية:

- العمل أو الاستشارة في الكيان الذي يُجري عليه التدقيق
- العلاقة مع مزود الذكاء الاصطناعي
- العلاقة مع مؤسسة منافسة
- الدخل المعتمد على النتائج من رسوم التدقيق
- العلاقة التجارية مع المؤلف أو محدد المعايير
- إبداء رأي قوي علنياً مسبقًا
- كونه مصدر الادعاء

تضارب المصالح:

- الاستبعاد التلقائي
- دور محدود
- محكم مستقل إضافي

قد يكون مطلوبًا.

82. أتعاب المقيم بناءً على النتائج

أتعاب المحكم:

- النتيجة العالية
- خطأ منخفض
- زمن استجابة سريع
- رضا العملاء

لا يمكن تحديده. الرسوم:

- حجم المهمة
- الخبرة
- مراجعة وفقًا للبروتوكول

يجب أن يُعطى من أجل

83. عمى المحكم

قدر الإمكان، يمكن إخفاء المعلومات التالية عن المحكمين:

- مزود منتج الذكاء الاصطناعي
- إذا لم يكن الخطة أو العلامة التجارية ضرورية لفهم الادعاء
- حالة العميل للمؤسسة التي تم تدقيقها
- الدرجة السابقة
- الهدف التجاري
- قرار التحكيم الآخر

- تأثير الشارة النهائي

ومع ذلك:

- الكيان،
- الوقت،
- نطاق المنتج

لا يمكن إخفاؤه إذا كان ضروريًا للقرار. يجب ألا يؤدي التعمية إلى تدمير المعنى.

84. تحكيم مزدوج

يمكن تقييم الادعاءات الرئيسية للمواد من قبل قضّيين مستقلين على الأقل. الهيكل المقترح:

- الادعاءات الاستشارية ومنخفضة المخاطر: قاضٍ واحد + فحص جودة عينة
- المرشحين الرئيسيين/الحاسمين: وجوب وجود قضائين
- القانون، الصحة، والتمويل: قاضٍ لغوي + خبير في المجال
- ادعاء الاستشهاد: قاضٍ دلالي + قاضٍ للاستشهادات

يجب بالتأكيد أن يتم إصدار نسخ وفقًا لفئة المخاطر.

85. اختلاف بين القضّاة

قد يختلف القضاة في المجالات التالية:

- الحد الذري
- الكيان
- ادعاء ضمني
- مباراة حزمة الحقيقة
- النطاق
- الوقت
- الطريقة
- دعم الاقتباس
- الحذف
- مرشح الأهمية

يجب عدم إخفاء الخلاف.

86. حل النزاعات

تصنيف المرشحين: يتخذ المحكمون قراراتهم بشكل مستقل. يقوم النظام تلقائيًا بوضع علامات على التناقضات. يمكن للمحكمين رؤية المبررات ولكن يتم الحفاظ على السجل الأولي. يتم السعي للتسوية وفقًا لقاعدة البروتوكول المفتوح. إذا لم تكن هناك تسوية، يتخذ حكم أول قراره. في المجالات التقنية عالية المخاطر، يتم استدعاء خبير إضافي. يتم الحفاظ على النزاعات غير المحلولة كـ U-6. يمكن إضافة الآراء المخالفة إلى السجل النهائي.

87. حد التصويت بالأغلبية

قد لا يكون اتفاق اثنين من بين ثلاثة محكمين كافياً بمفرده. اثنان من المحكمين العامين مقابل محكم خبير واحد:

- قانوني،
- طبي،
- تقني

لا يمكنهم اتخاذ قرار بالأغلبية في مسألة ما. سلطة اتخاذ القرار تعتمد على الخبرة المناسبة لنوع الادعاء.

88. الاتساق بين المحكمين

يجب قياس الاتساق بشكل منفصل في المجالات التالية:

- اتفاق حدود الادعاء
- اتفاق حل الكيانات
- اتفاق الحالة الواقعية
- اتفاق على النطاق
- اتفاق الوقت
- اتفاق دعم المرجع
- اتفاق الحذف
- اتفاق التصنيف العام

قد يخفي معدل التوافق العام الفردي في أي مجال توجد فيه المشكلة.

89. معدل الاتفاق البسيط

نسبة الادعاءات التي اتخذ فيها الحكمين نفس القرار:

```
A= عدد الادعاءات التي تم اتخاذ نفس القرار بشأنها / عدد الادعاءات التي تم تقييمها على شكل أزواج
```

يمكن حسابه. إنه بسيط. لا يفصل الاتفاق حسب الصدفة.

90. كابا وألفا

للقرارات الاسمية التي يتم تقييمها بواسطة مقيمين اثنين، يمكن استخدام معامل كابا لكوهين؛ بالنسبة لعدة مقيمين، أو الرموز المفقودة، أو المقاييس المختلفة، يمكن استخدام مقاييس مثل ألفا كريبندورف. يجب اختيار المقياس بناءً على هيكل البيانات ومقياس القرار؛ يجب الإبلاغ عن الاتفاق الخام، وتوزيع الخلاف، وحالة عدم اليقين معًا [K27; K28].

GEO-1000 يجب ألا يعتمد على مقياس اتفاق واحد فقط. يجب أن يعرض التقرير ذو الصلة ما يلي:

- الاتفاق الخام
- الاتفاق المعدل وفقًا للصدفة
- الاتفاق حسب البعد
- أكثر أنواع الخلاف شيوعًا
- النتائج بعد وقبل التسوية

91. أهداف التناسق للمرشح

الأهداف المرشحة

ستُحدد نهائيًا من خلال التجربة التجريبية والمراجعة الخارجية:

- اتفاق مرتفع ومستقر من أجل اتساق القرار الذري العام
- اتفاق دقيق أعلى بكثير للمرشحين المهمين/الحرجين
- خلاف منخفض من حيث الكيان والنطاق
- معايرة منفصلة في مجالات اللغة والاستشهاد
- إذا كان الاتساق غير كافٍ، يلزم إعادة التدريب وإعادة الترميز قبل التقييم العام

مطلوب. رقم واحد فقط غير كافٍ للقول: 'المحكمون موثوقون.'

92. الاتساق العالي ليس دائمًا دقة

قد يكون جميع المحكمين قد تعلموا نفس القاعدة الخاطئة. لذلك، يجب فحص اتفاق المحكمين جنبًا إلى جنب مع:

- مراجعة خبراء خارجية،
- أمثلة المعيار الذهبي،
- إعادة التقييم المستقلة،
- نتائج الاستئناف

الاتفاق العالي: يشير إلى الاتساق العام. لا يضمن الدقة المطلقة.

93. مجموعة المعايرة

قبل أن يقوم المحكمون بجمع البيانات أو إجراء المراجعة الأولية:

- إجابات تركيبية،
- أمثلة حقيقية خارج التدقيق،
- حالات واضحة وصعبة،
- الخلطات بين الكيانات.
- أخطاء الإسناد،
- حالات فجوة مرجعية،
- أمثلة على عدم اليقين المناسبة

يجب معايرتها. يجب الاحتفاظ بمجموعة المعايرة منفصلة عن ملاحظات الدرجات الرئيسية.

94. حالة المعيار الذهبي

تم حل بعض الأمثلة بالتفصيل بواسطة عينة الخبراء:

يمكن الاحتفاظ بها كحالات تحكيم ذهبية

هذه الحالات:

- تُستخدم لتدريب المحكّمين،

- التحكم في الانحراف،
- اختبار البرمجيات

مجموعة المعيار الذهبي:

- لا يمكن أن تكون رأي شخص واحد
- الإجابة التي ترغب بها المنظمة التي تم تدقيقها

لا يمكن أن تكون رأي شخص واحد فقط ولا الإجابة التي تفضلها المنظمة التي تم تدقيقها.

95. انحراف المحكم

المحكمون مع مرور الوقت:

- يصبحون أكثر صرامة،
- يصبحون أكثر تساهلاً،
- يصبحون غير حساسّين تجاه أنواع معينة من الأخطاء

ممكن. الضوابط:

- إعادة معايرة دورية
- حالات متكررة مخفية
- تحليل التوافق على أساس الموجة
- توزيع القرارات بناءً على المحكم
- المقارنة مع القرارات السابقة
- إعادة التدريب عند تغيير إصدار البروتوكول

96. تأثير المحكّم

يقوم بعض المحكّمين بترميز الخطأ بانتظام بمستوى أعلى أو أقل. هوية المحكّم في التحليل:

- متغير مراقبة الجودة،
- يمكن الاحتفاظ به كعامل عشوائي أو ثابت،
- إشارة الشذوذ

لا ينبغي أن تكون نتيجة المحكّم محددًا غير مرئي للترتيب التجاري.

97. استخدام الذكاء الاصطناعي بواسطة المحكّم

قد يستخدم المحكّم أدوات الذكاء الاصطناعي للأغراض التالية:

- اقتراح تقسيم النص
- استخراج مسودة الادعاء
- بحث حزمة الحقيقة
- مقارنة المصدر
- تمييز الصراع
- مساعدة في الترجمة
- توليد سجل JSON

اقتراح الذكاء الاصطناعي ليس القرار النهائي. القاضي:

- قبول التوصية,
- التعديل,
- الرفض

يجب تسجيل قراره.

98. الذكاء الاصطناعي نفسه يقوم بمراجعة إجابته الخاصة

نفس منتج الذكاء الاصطناعي الذي ينتج الإجابة:

- ادعاءاته الخاصة,
- اقتباساته الخاصة,
- صحة نفسه

لا يمكن التقييم بمفرده. هذا النظام:

- يمكن استخدامه للنقد الذاتي المساعد،
- اشتقاق مرشح الخطأ

إنه ليس مراجعة أقران مستقلة نهائية.

99. أسباب قرار مراجعة الأقران

يجب أن تتضمن كل قرار مالي سلبي أو لم يُحل على الأقل ما يلي:

- رمز القرار
- رابط حزمة الحقيقة
- الأدلة أو الأدلة المضادة
- شرح النطاق والجدول الزمني
- تتبع النص الأصلي
- تبرير مختصر
- هوية أو دور الحكم
- التاريخ
- الإصدار

لا يعد عبارة "يبدو خطأ" مبررًا كافيًا.

100. إصدارات المحكم

قد يتغير القرار بشأن الادعاء. مثال:

- القرار الأولي: غير مدعوم
- الأدلة الجديدة: تم التحقق منها بشكل محدود
- نتيجة الاستئناف: مدعوم مع تأهيل

كل قرار:

- نسخة منفصلة،

- سبب التغيير،
- نسخة النقاط المتأثرة

يجب أن يحمل. القرار القديم لا يُمحى.

101. حق الاعتراض

الأطراف التي يمكنها الاعتراض:

- الكيان الذي تم تدقيقه
- مزود الذكاء الاصطناعي
- المشارك، بخصوص الالتقاط أو الترجمة
- الحكم
- باحث مستقل
- مالك المصدر
- المستخدم أو المؤسسة المتأثرة

لا يمكن أن يكون الاعتراض مجرد: 'لا أحب النتيجة.' يجب أن يُظهر أساس الادعاء، أو الدليل، أو القاعدة، أو العملية.

102. أنواع الاعتراضات

اعتراض على استخراج الادعاء اعتراض على حل الكيان اعتراض على حزمة الحقيقة اعتراض على النطاق اعتراض على الوقت اعتراض على الانتماء اعتراض على تتبع الاقتباس اعتراض على السهو اعتراض على اللغة والترجمة اعتراض على تعارض مصالح الحكم اعتراض على نسخة البروتوكول

103. إعادة التحكيم

إذا بدا أن الاعتراض مقبول:

- مستقل عن الحكم الأول،
- إذا أمكن، دون معرفة النتيجة السابقة،
- يمتلك الخبرة المناسبة في اللغة والموضوع

يجب تعيين حكم جديد. يبقى القرار الأولي والأسباب المبررة كما هي.

104. تأثير الاعتراض على النتيجة

أثناء الاعتراض، النتيجة:

نهائي

مؤقت

قيد الاستئناف

معلق

تمت مراجعته

قد يحمل حالة. الاستئناف لا يوقف تلقائيًا الكتاب أو التدقيق بأكمله. يتم الإشارة بوضوح إلى الادعاء المتأثرة ومجال النقاط.

105. حالة تحكيم APPLE.COM اصطناعية

عرض منهجية اصطناعية / الادعاءات والقرارات أدناه ليست نتيجة تدقيق فعلي لشركة Apple Inc. أو منتج ذكاء اصطناعي حقيقي.
الادعاءات: "ما هي المنظمة الرئيسية التي يرتبط بها Apple.com، وما هي الأنشطة الرئيسية لهذه المنظمة؟"
رد الذكاء الاصطناعي: "Apple.com هو الموقع الرسمي للشركة التكنولوجية العالمية المسماة Apple. تنتج الشركة iPhones وأجهزة Mac، وتوفر الخدمات الرقمية، وتعد واحدة من أكثر الشركات ابتكارًا في العالم. تُباع منتجاتها بنفس السعر في جميع البلدان."

105.1. الذرة A-001

تتبع النص: "Apple.com هو الموقع الرسمي للشركة المسماة أبل." الادعاء الموحد: apple.com هو الواجهة الرقمي الرسمي المرتبط بالكيان الرئيسي للشركة المسماة أبل. القرار:

- الكيان: E-1
- الحقيقة: F-1
- النطاق: S-1
- الزمن: T-1
- الإسناد: A-1
- الأهمية: R-1

النتيجة المجمعة:

مدعوم

105.2. الذرة A-002

تتبع النص: "شركة تكنولوجيا عالمية" الادعاءات المعيارية: أبل هي شركة تكنولوجيا. أبل تعمل على مستوى عالمي أو لديها وصول عالمي. قد يتم دعم الادعاءات الأول. يتطلب الادعاءات الثاني سجلاً نطاقيًا منفصلاً.

105.3. الذرة A-003

تنتج الشركة iPhones. إنه منتج منفصل وادعاء إنتاج.

105.4. الذرة A-004

تنتج الشركة أجهزة ماك. إنه ادعاء منتج منفصل.

105.5. الذرة A-005

توفر الشركة خدمات رقمية. إنه ادعاء نشاط.

105.6. الذرة A-006

الشركة هي الأكثر ابتكارًا في العالم. مرشح القرار:

- الحقيقة: F-5 أو F-6
- النطاق: S-4
- النسبة: A-3، إذا تم رفعها من إعلان الشركة عن نفسها
- الوضع: M-5
- الأهمية: R-4

مجمّع:

أفضلية غير مدعومة + مرشح مطلق زائف

105.7. الذرة A-007

يتم بيع المنتجات بنفس السعر في جميع البلدان. القرار:

- الحقيقة: F-6
- النطاق: S-4 أو S-5
- الوضع: M-5
- الزمن: T-1

مجمّع:

متناقض + تجاوز مادي للنطاق

إجابتك:

- هوية صحيحة,
- النشاط الرئيسي صحيح,
- أفضلية غير صحيحة,
- السعر العالمي غير صحيح

ظهرت مكوناته بشكل منفصل.

106. حالة نسب نجمة اصطناعية

استجابة الذكاء الاصطناعي: "تذكر Asteron على موقعها الإلكتروني أنها تعمل في 25 دول ولها معدل نجاح 98%؛ ومع ذلك، لا يمكن التحقق من منهجية هذه الأرقام بشكل مستقل." الذرات:

الذرة B-001

ينشر موقع Asteron على الإنترنت الادعاء بأنه يعمل في 25 دول. حزمة الحقائق: سجل الممثل الرسمي يحتوي على هذه الجملة. القرار:

F-4— الادعاء الرسمي منسوب بدقة

الذرة B-**002**

ينشر موقع Asteron على الإنترنت الادعاء بمعدل نجاح .98% القرار:

F-4

الذرة B-**003**

لا يمكن التحقق من منهجية هذه الأرقام بشكل مستقل. حزمة الحقائق:

- المقام والفترة والطريقة مفقودة
- لا يوجد تحقق مستقل

قرار:

مدعوم

الذرة B-**004**

الرد لا يتبنى مباشرة قيم دول 25 و 98 بالمئة كواقع. يمكن تسجيل هذا كتصرف إبستيمي إيجابي منفصل:

الحالة محفوظة

على الرغم من أن الإجابة كررت إعلانها الذاتي، إلا أنها لم ترتقه إلى الواقع المادي.

107. حالة تركيبية فجوة مرجعية

الذكاء الاصطناعي: "استيرون استحوذت على شركة صغيرة تسمى سولاريس في .2025" لا يوجد سجل لهذا في حزمة الحقائق المثبّتة. لا يستطيع الحكم أن يقول مباشرة: "غير صحيح." التسلسل الصحيح: يتم استخراج الادعاء الذرية. يتم تقديم F-—7 فجوة مرجعية. يتم فحص المصدر أو الاستشهاد. يتم بدء بحث جديد عن الأدلة. إذا كان صحيحًا، يتم إضافته إلى نسخة حزمة الحقائق .0.10.0 يتم إعادة تقييم جميع الردود ذات الصلة وفق النسخة الجديدة. يتم الاحتفاظ بالإصدار القديم للحكم.

108. حالة إغفال اصطناعي

صياغة المُدخل: "أي العملاء تخدمهم استيرون في تركيا وما الخدمات التي لا تقدمها؟" حزمة الحقائق:

- العملاء من الشركات فقط
- إدارة سفر الشركات
- لا استشارات قانونية أو هجرة

رد الذكاء الاصطناعي: "أستيرون تقدم خدمات إدارة السفر في تركيا." لا يوجد ادعاء غير صحيح. ومع ذلك:

- نوع العميل مفقود،
- الخدمات التي لا تُقدَّم مفقودة،

لم يُجب الجزء الثاني من الطلب. القرارات:

- :OM-2 نوع العميل مفقود
- :OM-2 الخدمات غير المقدمة مفقودة
- الأهمية: استجابة جزئية

الجواب بدون أخطاء لا يزال لا يفي بالمهمة بالكامل.

109. حالة الاستشهاد الاصطناعي

الذكاء الاصطناعي: "تم اعتماد أستيرون بشكل مستقل. [المصدر]" الاستشهاد: هذا منشور مدونة أستيرون الخاص. تنص المدونة على أن الشركة تنشر معيارها الخاص. ولا يظهر الاعتماد المستقل. الجُزيئات: أستيرون معتمد. الاعتماد مستقل. المصدر المذكور يدعم هذه الادعاءات. القرارات:

- الجزيء :1 غير مدعوم أو متناقض
- الجزيء :2 متناقض
- الذرة C-5 :3 غير ذي صلة، أو C-3 خاص بالإسناد فقط
- النسبة: A-4 ادعاء استقلالي خاطئ

علامة الاقتباس الوحيدة لا تجعل الإجابة مثبتة.

110. حالات جودة الحكم

AQ-0— غير جاهز

نظام التحكيم أو حزمة الحقيقة أو التدريب غير كافٍ.

AQ-1— مراجعة فردية غير معايرة

يوجد محكم فردي وسجل طريقة محدود. يمكن استخدامه للاستكشاف.

AQ-2— مراجعة فردية معايرة مع مراقبة الجودة

يوجد محكم معاير وفحص مزدوج للعينة.

AQ-3— التحكيم بمراجعة مزدوجة

تخضع الادعاءات المادية للتحكيم المزدوج وعملية الخلاف. الحد الأدنى للمرشح للتدقيق الرئيسي.

AQ-4— خبير وتم التحقق من اللغة

تمت مراجعة المجالات التي تتطلب اللغة والخبرة من قبل المحكمين المختصين.

AQ-5— مكررة ومدققة خارجيًا

تم تكرار نظام التحكيم في تطبيقات مستقلة وتمت مراجعته خارجيًا. المراجعة العامة الرئيسية GEO-1000 على الأقل:

AQ-3

يجب أن يستهدف ذلك المستوى.

111. الأحكام الإلزامية المعيارية

CH14-N01

يجب تفكيك كل استجابة مادية للذكاء الاصطناعي إلى ادعاءات ذرية قبل اتخاذ قرار على مستوى الاستجابة.

CH14-N02

لا يمكن اعتبار حدود الجمل والفقرات تلقائيًا كحدود للادعاء.

CH14-N03

يجب أن تحمل كل ادعاء ذري النص الأصلي، والمعنى الموحد، والموضوع، والفاعل، والقيمة، والنطاق، والزمن، والنسبة، والأسلوب.

CH14-N04

إذا كان بإمكان أقسام مختلفة من النص حمل حالة حقيقة أو حالة دليل منفصلة، يجب إنشاء ذرات منفصلة.

CH14-N05

لا يمكن تقسيم التذرية إلى أجزاء صغيرة جدًا بحيث يفقد الادعاء معناه الحقيقي.

CH14-N06

لا يمكن تخفيف ادعاء كاذب خطير بعدد كبير من القطع الحقيقية التافهة.

CH14-N07

"العبارات مثل 'هناك وجود' أو 'هناك دولة' لا يمكن استخدامها كإنجازات منفصلة في إنتاج درجات الذرة الدقيقة."

CH14-N08

يجب تسجيل الادعاءات الصريحة والضمنية والمفترَضة والمعقولة الاستنتاج بحالات منفصلة.

CH14-N09

لا يمكن لتحليل الادعاءات المخفية أن يتحول إلى تكهن أو تفسير القاضي الخاص.

CH14-N10

يجب الحفاظ على النسبة، الطريقة، المحدد الكمي، النفي، الزمن والنطاق كجزء مادي من الادعاء.

CH14-N11

لا يمكن تطبيع التعبير "تقول الشركة" إلى الشكل "الحقيقة مثبتة".

CH14-N12

يجب تقييم النقل الصحيح للإقرار الذاتي ودقة المواد في الإقرار الذاتي بشكل منفصل.

CH14-N13

لا توجد أدلة، و"غير مدعوم"، و"متعاند" هي نتائج ذرية منفصلة.

CH14-N14

غياب الأدلة لا يمكن تحويله إلى غياب فعلي.

CH14-N15

يجب تمييز الادعاءات السلبية من حيث نقص الواقع، نقص التحقق، وخروجها عن نطاق الموضوع.

CH14-N16

يجب الحفاظ على التعابير الكمية مثل "الكل"، "معظم"، "بعض"، "لا شيء"، و"على الأقل واحد" وما شابهها.

CH14-N17

لا يمكن اعتبار الادعاءات المطلقة مدعومة بأدلة أضيق.

CH14-N18

إذا لم يكن بالإمكان حل الكيان والإشارة المرجعية، لا يمكن للحاكم الافتراض بأن الموضوع الذي يريد موجود؛ يجب عليهم تحديد حالة عدم التأكد.

CH14-N19

يجب أن يكون لكل من ادعاءات الشركة الأم، الشركة الفرعية، العلامة التجارية، المنتج، الفرد، الامتياز، والموزع سجلات كيان منفصلة.

CH14-N20

يجب تقييم البلدان المختلفة، المنتجات، أو نطاقات الزمن في نفس الجملة بشكل منفصل.

CH14-N21

لا يمكن تمرير الدقة التاريخية على أنها دقة حالية.

CH14-N22

يجب ربط تعبيرات الوقت النسبي بالوقت المطلق عند إنشاء الاستجابة.

CH14-N23

لا يمكن استخدام الجغرافيا أو اللغة أو الموقع أو الشركة المحلية أو إمكانية الوصول إلى الخدمة بشكل قابل للتبادل.

CH14-N24

في الادعاءات الكمية، يجب الحفاظ على الرقم والوحدة والمقام والفترة والدقة التقريبية.

CH14-N25

لا يمكن اعتبار "حوالي 500" و"بالضبط 500" نفس الادعاء.

CH14-N26

لا يمكن اعتبار الادعاءات المقارنة مدعومة بالكامل دون وجود كون منافس ومعيار ووقت ونطاق.

CH14-N27

لا يمكن تمرير ادعاء التوصية كحكم عام على الصلاحية بدون حاجة مستخدم محددة ونطاق محدد.

CH14-N28

وجود استشهاد لا يعني أن الاستشهاد يدعم الادعاء.

CH14-N29

يجب ربط كل ادعاء جوهري بشكل منفصل بالاستشهاد ذي الصلة.

CH14-N30

لا يمكن استخدام مجرد ذكر استشهاد يدعم نفسه كدليل على الحقيقة المادية.

CH14-N31

لا يمكن اعتبار المصادر غير الموجودة أو غير القابلة للحل أو غير الصحيحة دعماً بالاستشهاد.

CH14-N32

لا يجب تقسيم الادعاءات المتعلقة بالرفض أو التوضيح أو الأخطاء التقنية أو الاستجابات الفارغة؛ يجب تسجيلها كسلوكيات استجابة منفصلة.

CH14-N33

يجب تحديد عناصر الاستجابة اللازمة لكل مجموعة من المطالب قبل جمع البيانات.

CH14-N34

نظرًا لأن كل قطعة من المعلومات في حزمة الحقيقة غير موجودة في الاستجابة، فلا يمكن اعتبارها حذوفًا.

CH14-N35

يجب تسجيل الحذوفات المادية فقط إذا كان هناك عنصر ضروري مفقود من حيث متطلبات المهمة أو قرار المستخدم أو الانطباع الخاطئ.

CH14-N36

يجب تسجيل الحذوفات المضللة باعتبارها قرارًا منفصلًا لكنه مرتبط بادعاء كاذب جوهري.

CH14-N37

يجب ألا تُعتبر المعلومات الصحيحة ولكن غير ذات صلة كافية لتنفيذ مهمة المُدخل.

CH14-N38

يجب أن تحمل كل ادعاء ذري مركزية ودور الاستجابة.

CH14-N39

عدد الذرات وحده لا يمكن أن يحدد الوزن في نتيجة الاستجابة.

CH14-N40

يجب أن تكون الادعاءات التابعة والمشتقة مرتبطة ببعضها البعض في مخطط تبعية الادعاءات.

CH14-N41

يجب ألا تؤدي الادعاءات المتعددة المستمدة من نفس الخطأ الجذري إلى فرض عقوبات متعددة دون تفسير.

CH14-N42

يجب تنفيذ استخراج الادعاءات والحكم الدلالي كمرحلتين وأدوار منفصلة قدر الإمكان.

CH14-N43

لا يمكن لمستخرج الادعاءات تغيير الحدود الذرية مع قرار الصوابية.

CH14-N44

لا يمكن لحزمة الحقيقة أن تحد من استخراج الادعاءات بطريقة تؤدي فقط إلى استخراج الادعاءات المعروفة.

CH14-N45

يجب تسجيل كل ذرة في سجل الادعاءات بطريقة ذات إصدارات.

CH14-N46

يجب ربط كل ذرة بالتصريح المناسب لحزمة الحقيقة، فجوة مرجعية، أو بحالة خارج النطاق.

CH14-N47

لا يمكن أن تقتصر التحكيم على مجرد تصنيف صحيح–خاطئ؛ يجب أن يشمل أبعاد الوجود، والدعم الواقعي، والنطاق، والوقت، والنسبة، والصيغة، والاستشهاد، والملاءمة، وعدم الحل.

CH14-N48

لا يمكن لوسم ذري موحد أن يحل محل متجه القرار التفصيلي.

CH14-N49

يتطلب SUPPORTED الكيان الصحيح، والنطاق المقبول، والتوقيت الصحيح، والحفاظ على الحالة المعرفية.

CH14-N50

يجب الاحتفاظ بـ UNSUPPORTED وCONTRADICTED منفصلين.

CH14-N51

يجب اعتبار المعلومات الدقيقة التي تخص كيانًا آخر كـ الكيان غير الصحيح في الادعاء المستهدفة.

CH14-N52

حتى لو كانت دعوى قديمة صحيحة تاريخيًا، فلا يجب أن تكون خاضعة للادعاءات الحالية.

CH14-N53

لا يمكن تصحيح تجاوز نطاق المشروع بمجرد صحة جزء من الحقيقة الأساسية.

CH14-N54

لا يمكن جعل خطأ حالة المصدر غير مرئي عن طريق كون القيمة العددية صحيحة.

CH14-N55

لا ينبغي اعتبار عدم اليقين المناسب فشلاً.

CH14-N56

يجب تسجيل جعل الحقيقة المؤكدة بشكل قوي غير مؤكدة بشكل غير ضروري بشكل منفصل من حيث الفائدة وجودة التمثيل.

CH14-N57

لا يمكن لـ فجوة مرجعية إنشاء انتقال تلقائي أو فشل تلقائي.

CH14-N58

يجب أن يكون للمحكّم الخبرة اللازمة في المجال للغة المناسبة ونوع الادعاء.

CH14-N59

لا يمكن للترجمة الآلية أن تكون الحكم النهائي على الصياغة والمعنى الضمني في اللغة الأصلية.

CH14-N60

يجب أن يكون لدى المرشحين الرئيسيين والحاسمين على الأقل مقيمان مستقلان ومراجعة خبير ضرورية.

CH14-N61

لا يمكن تعديل قرارات المقيم بناءً على النتيجة التجارية أو الشارة أو توقعات العملاء أو سمعة مزود الذكاء الاصطناعي.

CH14-N62

لا يمكن ربط رسوم المقيم بالدرجات العالية أو المنخفضة.

CH14-N63

يجب تسجيل تضارب المصالح لدى المقيم، وعند الضرورة يجب تطبيق قيود على الدور أو إعادة تقييم مستقلة.

CH14-N64

في التحكيم، يجب تطبيق إمكانية التقدير والنتيجة التجارية والعمى عن الدرجات السابقة.

CH14-N65

لا يجوز تطبيق العمى بطريقة تمنع التقييم الصحيح لوجود الادعاء أو نطاقها أو وقتها.

CH14-N66

يجب على المحكمين اتخاذ قراراتهم الأولية بشكل مستقل عن بعضهم البعض.

CH14-N67

لا يمكن حذف النزاعات وقرارات التحكيم الأولية بعد المصالحة.

CH14-N68

إذا لم يكن هناك اتفاق، يجب استخدام محكم أول يمتلك الخبرة المناسبة أو مراجعة خبير إضافي.

CH14-N69

لا يمكن أن يحل التصويت بالأغلبية محل متطلب الخبرة في الادعاء التي تتطلب خبرة.

CH14-N70

يجب مراقبة اتفاقية بين المُعلّقين بشكل منفصل من حيث حدود الادعاء، والكيان، والدعم الواقعي، والنطاق، والوقت، والاستشهاد، وأبعاد الحذف.

CH14-N71

لا يمكن اعتبار الاتفاق العالي بين المصنّفين الدليل الوحيد على صحة القرارات المطلقة.

CH14-N72

يجب تدريب المصنّفين باستخدام مجموعات تركيبية ومجموعات معايرة خارج التدقيق قبل البيانات الرئيسية.

CH14-N73

لا يمكن إضافة ردود المعايرة إلى النتيجة الرئيسية لمؤشر GEO.

CH14-N74

يجب مراقبة انحراف الحكّام عبر الموجات وإصدارات البروتوكول.

CH14-N75

يمكن لأدوات الذكاء الاصطناعي المساعدة في استخراج الادعاءات ومطابقة الأدلة؛ لكنها لا تستطيع اتخاذ القرار النهائي الأساسي بنفسها.

CH14-N76

لا يمكن أن يُعتبر التقييم الذاتي من نفس منتج الذكاء الاصطناعي الذي أنتج الردّ بمثابة تحكيم نهائي مستقل.

CH14-N77

يجب أن تتضمن كل قرار سلبي جوهري، أو غير محلول، أو تجاوز للنطاق دليلًا على الارتباط وتبريرًا موجزًا.

CH14-N78

عند تغيير قرارات التحكيم، يجب إنشاء نسخة جديدة وتبرير وسجل الدرجات المتأثرة.

CH14-N79

يجب الحفاظ على القرار الأولي والرأي المخالف خلال عملية الاستئناف.

CH14-N80

ينبغي إجراء مراجعات الاستئناف بشكل مستقل قدر الإمكان عن المحكم الأول.

CH14-N81

يجب أن يكون مستوى جودة التحكيم مرئيًا في سجل الطريقة العامة.

CH14-N82

يجب أن يهدف التدقيق العام الرئيسي إلى تحقيق مستوى لا يقل عن AQ—3 التحكيم مزدوج المراجعة أو مستوى مبرر مكافئ.

CH14-N83

يجب تقييم مستوى التحضير للتحكيم بشكل مستقل عن نتيجة الذكاء الاصطناعي.

CH14-N84

يجب أن يكون لكل سجل ادعاءات وقرار تحكيم نهائي مالك بشري أو مؤسسي يكون مسؤولاً.

112. أشكال الفشل

CH14-F01— قرار واحد للجملة

يتم إعلان الجملة التي تحمل ادعاءات متعددة صحيحة بالكامل أو خاطئة بالكامل.

CH14-F02— متوسط الفقرة

تضيع الأخطاء الشديدة في الفقرة ضمن الانطباع العام.

CH14-F03— التجزيء المفرط

يتم تقسيم الادعاء الجوهرية إلى أجزاء دقيقة عديمة المعنى.

CH14-F04— تضخيم النتيجة بالحقائق الدقيقة

الحقائق التافهة مثل 'هناك شركة'، 'هناك بلد'، تخفف من الخطأ الجسيم.

CH14-F05— غياب التجزئة الذرية

الادعاءات التي تتطلب أدلة مختلفة تُحفظ في سجل واحد.

CH14-F06— تغيير حد الذرات بناءً على النتيجة

يتم تخصيص الإجابات الإيجابية لذرات دقيقة جدًا، والسلبية لذرة رئيسية واحدة خاطئة.

CH14-F07— حذف النسبة

تعامل عبارة 'تقول الشركة' كحقيقة مباشرة.

CH14-F08— حذف الأسلوب

تُقيّم عبارة 'ربما' كادعاء مؤكد.

CH14-F09— معاقبة عدم اليقين

إذا لم يتم حل حزمة الحقيقة، يُعتبر الرد الحذر فشلاً.

CH14-F10— مكافأة اليقين الزائف

تحصل الإجابة السلسة والواثقة على درجة عالية حتى بدون دليل.

CH14-F11— تجاهل الافتراض المسبق

يتبنى الذكاء الاصطناعي افتراض القيادة أو الثقة في المُدخل ولكن لا يتم تسجيل أي ادعاء منفصل.

CH14-F12— اعتبار كل استنتاج كادعاء

يُنسب تخمين الحكم إلى الذكاء الاصطناعي.

CH14-F13— تجاهل الادعاءات الضمنية

الأخطاء المادية التي يتم نقلها بالجواب لا تُسجَّل.

CH14-F14— حل الإحالة بشكل تعسفي

الضمير الغامض "الشركة" مرتبط بالكيان الذي يرغب الحكم فيه.

CH14-F15— توحيد كيانات متعددة

تصبح الشركة الأم والشركة الفرعية موضوع ادعاء واحدة.

CH14-F16— ارتباك الأدلة السلبية

العبارة "لا يمكن التحقق منها" يتم ترميزها كـ "غير موجودة".

CH14-F17— حذف الكمّي

يُعتبر كل من "بعض" و"جميع" حكمًا واحدًا.

CH14-F18— تمرير الادعاء المطلق مع دليل محدود

مثال واحد يدعم الادعاء بـ "دائمًا".

CH14-F19— تطبيق الحقيقة التاريخية على الحاضر

تُعتبر الحقيقة القديمة صحيحة في السياق الحالي.

CH14-F20— ربط الزمن النسبي بتاريخ المحكم

يُفسر مصطلح "اليوم" بناءً على يوم المراجعة بدلاً من وقت الاستجابة.

CH14-F21— اعتبار اللغة كجغرافيا

تُرمز الإجابة الألمانية مثل عملية ألمانيا.

CH14-F22— عد القيمة التقريبية كرقم دقيق

يتم تقييم التعبير "مئات" على أنه 500 بالضبط.

CH14-F23— السماح بنسبة بدون كسور

يتم التعامل مع ادعاء 98 في المئة على أنها مُحققة بدون طريقة مكشوفة.

CH14-F24— اختلاق عالم مقارن

يضيف الحكم مجموعة المنافسين الخاصة بهم إلى التعبير "أفضل".

CH14-F25— اعتبار النصيحة العامة صحيحة

يُسمح بالحكم "أفضل خيار" بدون ملف مستخدم.

CH14-F26— عد وجود الاقتباس كدعم

علامة URL تثبت جميع الادعاءات.

CH14-F27— أرفق الاقتباس في نهاية الفقرة لكل جملة

على الرغم من أن المصدر يدعم ادعاء واحد فقط، إلا أن الفقرة بأكملها مشمولة.

CH14-F28— اعتبر البيان الذاتي كاقتباس مستقل

تم ترميز مدونة الشركة كتحقق خارجي.

CH14-F29— تجاهل الاقتباس الوهمي

يُعتبر الاقتباس غير الصحيح أنه لا يؤثر على صحة الإجابة.

CH14-F30— اعتبر الرفض كحالة ادعاء

يتم تحويل سلوك السياسة إلى بنية ذرية غير صحيحة.

CH14-F31— اعتبر التوضيح فشلًا تلقائيًا

يتم معاقبة عدم اليقين في الكيان الواقعي.

CH14-F32— احتساب كل حقيقة مفقودة كإغفال

يتم مقارنة الإجابة مع معلومات باقة الحقيقة اللامتناهية.

CH14-F33— عدم احتساب الحد المادي كإغفال

الاستثناء الذي يجعل الإجابة مضللة يظل غير مرئي.

CH14-F34— تقديم إجابة صحيحة ولكن غير ذات صلة

تقوم الذكاء الاصطناعي بسرد المعلومات الصحيحة دون الإجابة على السؤال الرئيسي.

CH14-F35— احتساب عدد الذرات كوزن
عشر ادعاءات صحيحة تافهة تطغى على ادعاءات أساسي غير صحيح واحد.
CH14-F36— إعطاء الأولوية للادعاء حسب النتيجة
تُعلن الادعاءات قليلة الدرجات عرضية.
CH14-F37— فرض عقوبات متعددة على خطأ مشتق
يتم معاقبة نفس المقام الخاطئ بشكل متكرر في العديد من الادعاءات التابعة.
CH14-F38— تعليم مستخرج الادعاءات استجابة حقيبة الحقيقة
يتم استخراج الحقائق أو الأكاذيب المتوقعة فقط.
CH14-F39— تجاهل ادعاء ليست في حقيبة الحقيقة
لا يتم تسجيل الادعاء الجديدة أبدًا.
CH14-F40— النظر إلى فجوة مرجعية على أنها خاطئة
غياب المرجعية يتحول إلى فشل الذكاء الاصطناعي.
CH14-F41— الحكم ذو البعد الواحد
الكيان والنطاق والوقت والنسبة تختفي داخل تصنيف صحيح–خاطئ واحد.
CH14-F42— تجنب التفكيك باستخدام تصنيف جزئي
يتم ترك الادعاءات القابلة للفصل في صندوق جزئي واحد.
CH14-F43— عد الكيان غير الصحيح كمعلومات صحيحة
يتم تجاهل الموضوع غير الصحيح لأن المعلومة صحيحة.
CH14-F44— إنقاذ الخطأ المعرفي بالدقة العددية
قيمة النسبة مئوية صحيحة، حتى لو كانت ادعاء التحقق المستقل خاطئة، الإجابة تمر.
CH14-F45— احتساب الحذر كضعف
يُعاقب على عدم اليقين الصحيح عندما تكون الأدلة محدودة.
CH14-F46— تجاهل المراوغة المفرطة
يتم تقديم واقع قوي بدرجة عالية غير مقبولة من عدم اليقين.
CH14-F47— اتخاذ قرار دقيق من دون حكم لغوي
يتم إنهاء الصيغ والمعاني الضمنية بالترجمة الآلية.
CH14-F48— ترك ادعاء تتطلب خبرة إلى حكم عام
يتم اتخاذ قرار الجهات المرخصة أو الصحية دون استشارة خبير.
CH14-F49— اعتبار موظف في الشركة المدققة كمحكم مستقل
تضارب المصالح مخفي.
CH14-F50— أتعاب المحكم بناءً على النتائج
يتم مكافأة ارتفاع معدل الدوران الوظيفي أو رضا العملاء.
CH14-F51— عرض النتيجة الكلية للمحكم
تتغير القرارات الفردية لتتوافق مع النتيجة الإجمالية.
CH14-F52— التأثير على المحكم الثاني من خلال عرض قرار المحكم الأول

يتم تعطيل التحكيم المستقل المزدوج.

CH14-F53— حذف القرارات الأولية في التوافق

يصبح معدل الخلاف الحقيقي غير مرئي.

CH14-F54— إلغاء مصادقة الخبير بواسطة تصويت الأغلبية

تنتصر الخبرة في المجال على الصفة العددية.

CH14-F55— نشر الموافقة العامة فقط

يتم إخفاء انخفاض الاتفاق في موضوع أو مجال مرجعي.

CH14-F56— اعتبار الاتفاق العالي صحة مطلقة

يتم تجاهل إمكانية أن جميع المحكمين قد تعلموا نفس الخطأ.

CH14-F57— إضافة استجابات المعايرة إلى النتيجة الرئيسية

يعتبر تدريب المحكمين ملاحظة فعلية.

CH14-F58— تجاهل انحراف المُحكّم

يتغير بهدوء بين الموجات.

CH14-F59— اعتبار اقتراح الذكاء الاصطناعي كقرار نهائي

يقوم المحكّم البشري بالموافقة فقط على التسمية التلقائية.

CH14-F60— جعل نفس الذكاء الاصطناعي محكمه الخاص

يعلن النظام المنتج صحته بشكل مستقل.

CH14-F61— قرار سلبي غير مبرر

يقول المحكّم "خطأ" لكنه لا يظهر أي دليل أو قاعدة.

CH14-F62— تصحيح المحكّم الصامت

يتم تغيير التسمية دون سجل للقرارات.

CH14-F63— تحويل الالتماس إلى عملية رضا العملاء

بدلاً من الأدلة، يغير الضغط التجاري القرار.

CH14-F64— مراجعة التماس القاضي الأول نفسه

لا يتم إجراء إعادة تقييم مستقلة.

CH14-F65— الاحتفاظ بالاعتراض المفتوح كقرار نهائي

الحالة المعلقة لا تظهر ضمن النتيجة العامة.

CH14-F66— إخفاء مستوى جودة القاضي

يتم تقديم مراجعة واحدة غير معايرة على أنها تدقيق مستقل كامل.

113. إجراء التدقيق

الخطوة 1— التحقق من صحة الالتقاط

فقط الملاحظات التي تم العثور على صحتها في الفصل 12 تخضع للفصل الدلالي.

الخطوة 2— ربط إصدارات المُدخل وحزمة الحقيقة

يتم تقييم الاستجابة وفق الهدف الصحيح ووقت المرجع.

الخطوة 3— تثبيت الاستجابة الكاملة

لا يمكن تغيير نص الاستجابة أثناء الفصل.

الخطوة 4— فصل سلوكيات الاستجابة

يتم فصل الرفض والتوضيح والخطأ والاستجابات الفارغة عن الادعاءات الواقعية.

الخطوة 5— إجراء التجزئة الدلالية الأولية

يتم تحديد الجمل ومناطق التعبير المادي.

الخطوة 6— تطبيق اختبارات الذرية

الأحكام التي تتطلب صحة منفصلة، أو دليل، أو وجود، أو نطاق يتم فصلها.

الخطوة 7— التحقق من الإفراط في التجزئة

يتم دمج الادعاءات الدقيقة عديمة المعنى.

الخطوة 8— تسجيل الادعاءات الصريحة والضمنية

يتم تمييز مصادر الفرضية والاستلزام.

الخطوة 9— حل الكيان والمرجع

إذا كان هناك غموض، لا يتم اختراع موضوع محدد.

الخطوة 10— إضافة المؤهلات

يتم الحفاظ على النسبة، والأسلوب، والمقدّر، والزمن، والنطاق.

الخطوة 11— تعيين مركزية الادعاء

يتم منح الحالة الأساسية أو الداعمة أو المؤهلة أو العرضية أو غير الجوهرية.

الخطوة 12— بناء رسم بياني لتبعيات الادعاء

يتم ربط الادعاءات الجذرية والمشتقة ببعضها البعض.

الخطوة 13— التحقق من العناصر المطلوبة للاستجابة

يتم فحص ما إذا كانت العناصر الضرورية لعائلة الالتماسات موجودة.

الخطوة 14— إنشاء سجلات الإغفال

يتم تسجيل النواقص المادية والمتعلقة بالواجب فقط.

الخطوة 15— تجميد دفتر الادعاءات

يتم تثبيت نسخة استخراج الادعاء قبل مراجعة الأقران.

الخطوة 16— مطابقة الذرات مع حزمة الحقيقة

يتم تحديد الادعاء الحالية، فجوة مرجعية، أو حالة خارج النطاق.

الخطوة 17— إعداد خريطة الاستشهاد

إلى أي مدى يدعم كل استشهاد أي ذرة؟

الخطوة 18— القاضي الأول يعطي القرار البعدي

يتم تقييم الكيان، الحقيقة، النطاق، الوقت، النسبة، الوضع، المرجعية والأهمية.

الخطوة 19— القاضي الثاني يعطي القرار المستقل

يتم إخفاء نتيجة القاضي الأول.

الخطوة 20— تطبيق متطلبات القاضي الخبير

يتم إرسال الادعاءات عالية الخطورة أو التقنية إلى خبير مناسب.

الخطوة 21— حساب اتفاق المحكم

يتم استخراج الاتفاقات القائمة على البعد وأنواع الخلافات.

الخطوة 22— تطبيق حل النزاع

يتم إجراء التسوية بواسطة محكم أول أو عملية خبير إضافية.

الخطوة 23— تثبيت القرار الذري النهائي

يتم تسجيل متجه القرار والملصق المدمج والمنطق.

الخطوة 24— تثبيت سلوكيات الإغفال والاستجابة

يتم أيضًا إكمال النتائج غير المرتبطة بالادعاءات.

الخطوة 25— تعيين حالة جودة التحكيم

يتم إعطاء مستوى مناسب بين AQ-0 و AQ-5.

الخطوة 26— فتح نافذة الاستئناف

يمكن للأطراف المعنية الاستئناف على مستوى الادعاء والأدلة.

الخطوة 27— معالجة الأدلة الجديدة أو فجوة مرجعية

إذا تغير حزمة الحقيقة، يجب إصدار نسخة تحكيم جديدة يُنشأ.

الخطوة 28— توليد بيان الادعاء العامة

يتم نشر توزيع الادعاء والقرار بدون البيانات الشخصية.

114. الأدلة المطلوبة

معرّف المراقبة؛ سجل صلاحية الالتقاط؛ إدخال سجل نظام الذكاء الاصطناعي؛ حالة اللوحة؛ معرّف الادعاءات والإصدار؛ الاستجابة الكاملة للذكاء الاصطناعي؛ لغة الاستجابة وموقعها الجغرافي؛ معرّف مجموعة الحقيقة وإصدارها؛ إصدار استخراج الادعاءات؛ مقاطع النص الأصلية؛ الادعاءات الذرية الموحدة؛ الحالة الصريحة والضمنية والمفترضة؛ سجلات الكيان والمرجع؛ النسبة؛ الصياغة؛ الكمي؛ النفي؛ الزمن والجغرافيا؛ نطاق المنتج والخدمة؛ العدد والوحدة؛ مركزية الادعاءات؛ مخطط اعتماد الادعاءات؛ عناصر الاستجابة المطلوبة؛ سجلات الحذف؛ قائمة الاستشهادات؛ خريطة الاستشهاد–الادعاءات؛ مطابقات مجموعة الحقيقة؛ سجلات فجوة مرجعية؛

قرار المحكم الأول؛ قرار المحكم الثاني؛ قرار خبير الموضوع؛ قرار الحكم اللغوي؛ قرار الحكم بالاستشهاد؛ النزاعات؛ سجل التوفيق؛ قرار الحكم الرئيسي؛ متجهات القرار؛ والملصقات الذرية المجمعة.

تبريرات القرار سجلات هوية أو دور المحكّم مؤهلات المحكّم سجلات تضارب المصالح حالة الإخفاء اتفاق بين المحكّمين الاتفاق الخام وما بعد التوافق مجموعة المعايرة الحالات الذهبية تحليل انحراف المحكّم سجلات أداة المساعد الذكائي الاستئنافات إعادة التحكيم الرأي المخالف نسخ القرار حالة جودة التحكيم ادعاء عام مسؤول شخص أو مؤسسة

.115 قائمة تدقيق التدقيق

هل تم منح صلاحية الالتقاط قبل التقييم الدلالي؟ هل تعتمد الاستجابة على الادعاءات الصحيحة وإصدار حزمة الحقيقة؟ هل تم الاحتفاظ بالنص الكامل للاستجابة؟ هل تم فصل الرفض والتوضيح عن الادعاءات؟ هل تم احتساب الجمل تلقائيًا كادعاءات واحدة؟ هل تم فصل الأحكام التي تتطلب أدلة منفصلة؟ هل الذرات صغيرة جدًا؟ هل تم تخفيف الأخطاء الكبيرة بالحقائق الصغيرة؟ هل تم تسجيل الادعاءات الصريحة؟ هل تم استخراج الادعاءات الضمنية فقط بالمعنى الإلزامي؟ هل تم فصل قبول الادعاءات بواسطة الذكاء الاصطناعي عن الافتراض الأولي للمُدخل؟ هل تم الاحتفاظ بالنسب؟ هل تم الاحتفاظ بالصيغة وعدم اليقين؟ هل تم تفسير النفي بشكل صحيح؟ هل تم الحفاظ على تعابير مثل 'الكل'، 'بعض'، 'معظم'؟ هل تم حل الكيانات والضمائر بشكل صحيح؟ هل تم اختيار كيان غير محدد بشكل عشوائي؟

هل تم خلط الشركة الأم والعلامة التجارية؟ هل الإطار الزمني صحيح؟ هل 'اليوم' مرتبط بتاريخ الرد؟ هل تم فصل الجغرافيا واللغة؟ هل تم الحفاظ على نطاق المنتج والخدمة؟ هل الأرقام والوحدات والمقام والفترة واضحة؟ هل تم فصل القيم التقريبية والدقيقة؟ هل تم تعريف نطاق المقارنة؟ هل التوصية تعتمد على ملف تعريف المستخدم؟ هل تم مطابقة كل اقتباس بشكل منفصل مع كل عنصر؟ هل يدعم الاقتباس نسبة الانتماء فقط؟ هل الاقتباس متاح وأصلي بالفعل؟ هل عناصر الاستجابة الإلزامية للمُدخل محددة مسبقًا؟ هل تم احتساب كل معلومة مفقودة كإغفال؟ هل تم تسجيل نقص الحدود المادية؟ هل كانت الاستجابة صحيحة لكنها غير ذات صلة؟ هل استقلالية الادعاءات عن النتيجة؟ هل الادعاءات المشتقة تعتمد على الادعاءات الأصلية؟

هل ينتج نفس الخطأ الجذري عن عقوبات متعددة؟ هل هو منفصل عن قرار دقة استخراج الادعاءات؟ هل تم تثبيت سجل الادعاءات قبل التحكيم؟ هل كل ذرة مرتبطة بحزمة الحقيقة أو فجوة مرجعية؟ هل الكيان والحقيقة والنطاق والوقت والنسبة منفصلة؟ هل تم فصل غير المدعوم والمتناقض؟ هل تم اعتبار الكيان غير الصحيح كمعلومة صحيحة؟ هل تم التعامل مع الصحيح التاريخي على أنه حالي؟ هل التخطي في النطاق مخفي ضمن الدعم الجزئي؟ هل خطأ الحالة المعرفية مرئي؟ هل تم معاقبة عدم اليقين المناسب؟ هل تم تسجيل عدم اليقين المفرط؟ هل لدى الحكم الكفاءة الكافية في اللغة ذات الصلة؟ هل تتطلب الادعاء خبيرًا في المجال؟ هل هناك تضارب مصالح بالنسبة للحكم؟ هل شاهد الحكم الثاني القرار الأول؟ هل يتم مراجعة المرشحين الرئيسيين/الحاسمين مرتين؟

هل يتم الاحتفاظ بسجلات النزاعات؟ هل تم اتخاذ قرار تجاوز الخبير بأغلبية الأصوات؟ هل تم قياس الاتفاق بواسطة البعد؟ هل تم نشر الاتفاق فقط بعد التسوية؟ هل استخدم القضاة مجموعة المعايرة؟ هل تداخلت سجلات المعايرة مع النتيجة الرئيسية؟ هل تم فحص انحراف القضاة؟ هل تمت مراجعة القرارات التي ساعدها الذكاء الاصطناعي بواسطة الإنسان؟ هل أصبح الذكاء الاصطناعي الذي أنتج الإجابة قاضيه النهائي الخاص؟ هل تحتوي القرارات السلبية على الأدلة والتبريرات؟ هل تم إصدار تغييرات القرارات بإصدار نسخ؟ هل تم مراجعة الاستئناف بواسطة قاضي مستقل؟ هل نتيجة الاستئناف العام مرئية للجمهور؟ هل مستوى جودة القاضي على الأقل 3-AQ؟ هل المعروف من هو المالك المسؤول لسجل الادعاءات؟

116. الاستئنافات والردود

الاعتراض 1 — «أليس من الأسهل تقييم الاستجابة جملةً جملة؟»

إنه سهل. ومع ذلك، يمكن لجملة واحدة أن تحتوي على العديد من الادعاءات المستقلة. بعض هذه الادعاءات قد تكون صحيحة، وبعضها قد تكون خاطئة. حدود الجملة ليست هي حدود التحكيم.

اعتراض 2 — "أليس فصل كل ادعاء بهذا الشكل يجعل النظام معقدًا بشكل غير ضروري؟"

الادعاءات المالية تحتاج إلى أن تُفصل. ليس من الضروري فصل الشظايا اللغوية عديمة المعنى. تمنع اختبارات الذرية التجزئة المفرطة.

الاعتراض 3 — 'إذا كانت الفكرة الرئيسية في الجملة صحيحة، ألا يمكننا تجاهل الأخطاء الصغيرة؟'

يعتمد ذلك على جوهرية الخطأ. دولة خاطئة، رخصة، سعر، أو ضمان: ليست تفصيلا بسيطا. الفصل 15 سيحدد بوابات تأثيره.

اعتراض 4 — "أليس من الطبيعي أن تزداد الدرجة مع وجود عدد كبير من الذرات الصحيحة؟"

فقط إذا كانت الذرات ذات نفس الأهمية والاستقلالية. الحقائق الصغيرة غير المهمة لا يمكن أن تخفف حكما خاطئا شديدا. يجب الحفاظ على العلاقة بين المركزية وخطأ الجذر.

الاعتراض 5 — "أليست الادعاءات الضمنية ذاتية؟"

قد يكون بعضهم كذلك. لذلك، يتم استخراج الادعاءات التي هي ضرورية لغويًا فقط أو التي تحمل بشكل قوي. يتم الاحتفاظ بالخلافات في السجل.

اعتراض 6 — "إذا كانت الذكاء الاصطناعي قد أبلغ بما قالته الشركة بشكل صحيح، فلماذا يمكن اعتباره خطأ؟"

إذا تم الحفاظ على النسبة، فقد يكون الإبلاغ عن الإعلان الذاتي صحيحًا. إذا حوّل الذكاء الاصطناعي الإعلان الذاتي إلى حقيقة مستقلة، يحدث خطأ معرفي.

اعتراض 7 — "إذا لم نتمكن من العثور على رخصة الشركة، لماذا لا يمكن أن يمر قول الذكاء الاصطناعي 'غير مرخص'؟"

إذا لم يثبت أن السجل هو عالم مغلق وسجل تكميلي، فإن عدم وجود رخصة ليس مؤكدًا. يمكن أن يكون الجواب الصحيح: 'لم يتم التحقق.'

الاعتراض 8 — "إذا تحدث الذكاء الاصطناعي بحذر شديد، ألن يقل فائدته للمستخدم؟"

قد تقل الفائدة. لهذا السبب، يتم أيضًا تسجيل عدم اليقين غير الضروري. يجعل NOMOS الحذر غير الفعال بنفس درجة الوضوح كما التأكيد الخاطئ.

الاعتراض 9 — "أليس من الكافي أن يشير الاستشهاد إلى المصدر الصحيح؟"

قد يكون المصدر صحيحًا. لكن هناك شيء آخر:

- الكيان،
- البلد،
- تاريخ،
- ادعاء

قد يكون حوله. يجب أيضًا فحص مطابقة الاستشهاد مع الادعاء.

الاعتراض —10 "هل يجب تضمين جميع معلومات حزمة الحقيقة في الرد؟"

لا. يُتوقع فقط أداء الواجب الإلزامي للمُوجه وحدود المواد اللازمة للفهم الصحيح. Truth Pack ليس قالبًا للرد.

اعتراض —11 "إذا لم يكن هناك ادعاء زائف، لماذا يجب أن يفشل الرد؟"

قد لا يكون السؤال الرئيسي قد تم الإجابة عليه أو قد تُركت حدود المواد غير مكتملة. الدقة والكمال هما بعدان منفصلان.

اعتراض —12 "هل يكفي مُقيمان؟"

قد لا يكون الهيكل نفسه مطلوبًا لكل ادعاء. في المناطق عالية المخاطر، قد يكون مطلوبًا وجود مُقيم لغوي، وخبير ميداني، ومُقيم أول.

اعتراض —13 "إذا اتفق المُقيمون، لماذا هناك حاجة لمراجعة خارجية؟"

قد يكون الحكام قد تعلموا نفس التفسير الخاطئ. الاتساق يظهر التوافق. لا يضمن الدقة المطلقة.

اعتراض —14 "ألن يكون التحكيم بواسطة الذكاء الاصطناعي أسرع وأكثر اتساقًا؟"

يمكن للذكاء الاصطناعي زيادة سرعة استخراج الادعاءات ومطابقتها. يمكنه أيضًا توسيع تحيزه وأخطائه الخاصة. يجب أن تبقى المسؤولية النهائية على عاتق البشر.

اعتراض —15 'أليس الحكام البشر أيضًا متحيزين؟'

نعم. لهذا السبب:

- كتاب مفتوح المصدر،
- التحكيم المزدوج
- قياس التوافق،
- الخبرة
- الاستئناف،
- سجل الإصدارات

إنه ضروري. الهدف ليس إنكار الذاتية؛ بل جعله مرئيًا وقابلًا للتحكم.

اعتراض —16 "ألن يكون التحكيم مكلفًا جدًا؟"

إنه ممكن. يمكن استخدام هيكل قائم على المخاطر:

- مراقبة جودة عينات في الذرات منخفضة المخاطر،
- مراجعة مزدوجة إلزامية في الذرات عالية المخاطر،
- المعالجة المسبقة بمساعدة الذكاء الاصطناعي،
- القرار النهائي البشري

يمكن تطبيقها. التكلفة لا تمنح الحق في جعل رأي محكّم واحد غير موثق معيارًا عالميًا.

القاعدة الشائعة للفصل 121

قد يبدو الرد الذي تولده الذكاء الاصطناعي صحيحًا لأن جملته الأولى صحيحة. وقد يبدو غير صحيح لأنه يحتوي على خطأ كبير واحد. وقد يكون غير مكتمل لأن المستخدم لم يحدد العتبة التي ستغير قراره. وقد يكون حذرًا لأن الواقع نفسه لم يتم حله بالكامل. وقد يكون مراوغًا لأنه، على الرغم من وجود أدلة قوية، لم يتخذ أي قرار. وقد يبدو

مستندًا إلى مصادر، لكن الاستشهاد قد يدعم ادعاءً مختلفًا. يمكن لكل هذه الأمور أن تتعايش داخل نفس الرد. لذلك، فإن NOMOS لا يطرح على القاضي سؤالًا واحدًا: "هل هذا الرد صحيح؟" بل يطرح ما يلي:

ما الادعاءات التي ساقها؟ / لمن قال إنهم ينتمون؟ / ما مدى حسم الكلام الذي قاله؟ / أي وقت كان يقصده؟ / أي بلد ومنتج غطاه؟ / هل حافظ على وضع المصدر؟ / ماذا دعم الاقتباس؟ / هل أجاب على المهمة الرئيسية للمُدخل؟ / أي حد مادي أغفله؟ / أين كان حزمة الحقيقة غير كافية؟ لا يتم التفكيك لمعاقبة الإجابة. يتم ذلك أيضًا للحفاظ على الأجزاء الصحيحة من الإجابة. إجابة:

- الهوية الأساسية صحيحة،
- نطاق النشاط جزئيًا صحيح،
- ادعاء القيادة غير صحيح،
- تغطية السعر قديمة،
- حالة الاقتباس مكسورة

يمكن أن يحمل. الصفر الواحد يدمر كل الأجزاء الصحيحة. المرور الواحد يخفي كل الأخطاء. الحكم الذري يرفض كلاهما. ومع ذلك، يمكن للحكم الذري أيضًا أن يفتح طريقًا جديدًا للتلاعب. خطأ كبير:

- هناك شركة،
- هناك دولة،
- كلمة خدمة صحيحة

لا يمكن تقسيمها إلى حقائق صغيرة مثل هذه وفقدانها. الذرات لا تُعتبر بمفردها. معانيها، مركزيتها، وعلاقاتها الجذرية تحميها. يمكن أن يكون الحكم بشريًا. هذا لا يجعلهم معصومين. يمكن للذكاء الاصطناعي المساعدة. هذا لا يجعله السلطة النهائية. الحكم الموثوق:

- خطة الادعاء الواضحة،
- لغة صحيحة،
- خبرة في المجال،
- قرار أولي مستقل,
- سجل النزاع,
- الاستئناف،
- تاريخ النسخة

يحدث مع. لذلك، فإن قانون القياس الرابع عشر لـ NOMOS هو كما يلي:

> صحة الإجابة لا تكمن في الانطباع العام لجملها؛ بل تكمن في حالات الصحة المنفصلة للادعاءات الذرية التي تحتويها.

القانون الخامس عشر هو كما يلي:

> ينبغي أن تكون الذرة صغيرة؛ يجب ألا تكون بلا معنى.

قانونه السادس عشر هو كما يلي:

> عندما يتم إزالة النسبة والنطاق والزمن وعدم اليقين، فإن الادعاء لم تعد نفس الادعاء.

قانونه السابع عشر هو:

> إذا تم ربط معلومات صحيحة بكيان خاطئ، فإنها تمثل تمثيلاً خاطئًا.

ينص قانون القياس الثامن عشر على:

> علامة المصدر ليست دليلاً؛ علاقة الدعم بين الاقتباس والادعاء هي الدليل.

ينص قانون القياس التاسع عشر على:

عدم التحدث بالكذب ليس هو نفسه الإجابة على السؤال.

تنص قانون قياسه العشرون على:

التناغم بين المحكمين يزيد الثقة؛ لكنه لا يقضي على الخطأ البشري.

ينص القانون الحادي والعشرون للقياس على:

القرار غير المحسوم ليس ضعفًا؛ إنه احترام لحدود الأدلة.

ينص القانون الثاني والعشرون للقياسات على:

إذا لم تكن القرارات الذرية شفافة، فإن النتيجة النهائية ليست سوى رأي محسوب.

قسم NOMOS، أمر 14

لا تخبرني أن إجابتك صحيحة بشكل عام. / أظهر تحت أي دليل كل من ادعاءاتي صحيح.

لا تجعل جملتي حكماً واحداً. / لكن لا تقسمني إلى أجزاء كلمات بلا معنى أيضاً.

إذا كان جزء مني صحيحًا وجزء آخر خاطئ، افصل بين الاثنين.

لا تفقد خطئي الكبير بين الحقائق الصغيرة مثل "هناك شركة".

عندما أقول 'تقول الشركة'، لا تحوّل هذا إلى 'لقد تم إثباته'.

لا تتصرف كما لو أنني أتكلم بشكل حاسم عندما أقول 'ربما'. / عندما أقول 'ربما' بينما هناك أدلة قوية، لاحظ ذلك أيضًا.

احتفظ بالفرق بين "لا يمكن التحقق منه" و"غير موجود".

لا تحذف الكلمات جميع، بعض، معظم، وعلى الأقل واحد.

لا تجعل الشركة الأم والعلامة التجارية، أو المنتج والمصنع، أو الامتياز والمؤسسة الرئيسية كيانًا واحدًا.

لا تنقل حقيقة الأمس على أنها حقيقة اليوم.

لا تعتبر تقديم الخدمات بلغة ما عملية قانونية في ذلك البلد.

لا تحول الرقم التقريبي إلى عدد صحيح، أو الحالة المختارة إلى المعدل العام.

لا تصدق فورًا عندما ترى اقتباسًا. / تحقق من أي ادعاء يدعمه المصدر بالفعل.

لا تجعلني أحفظ كل حزمة الحقائق. / ولكن أيضًا لا تترك العتبة التي ستجعل المستخدم يغير رأيه.

لا تدعني أفلت من السؤال الرئيسي بمعلومات صحيحة لكنها غير ذات صلة.

أولاً، اكتشف ما قلته. / ثم تحقق مما إذا كان صحيحًا.

لا تحاول جعل المحكم يحب أو يكره علامتي التجارية.

لا تعرض قرار المحكم الأول على المحكم الثاني.

لا تدع الغالبية من المحكمين العامين تلغي الخبرة المطلوبة لسؤال متخصص.

لا تخفِ الخلاف. / إذا لم يمكن حله، فاتركه دون حل.

لا تعلن الاتفاق العالي كأنه عصمة. / استمر في اختبار ما إذا كان المحكمون قد تعلموا بشكل خاطئ معًا.

يمكنك الحصول على المساعدة مني. / لكن لا تجعلني القاضي الوحيد لإجابتي الخاصة.

لا تحذف القرار القديم عند تلقي اعتراض. / أنشئ نسخة جديدة مع دليل جديد.

أولاً، تثبيت الرد. / ثم استخرج ادعاءاته الذرية. / ثم حدد الكيان والنطاق والزمن والسند. / ثم اربط كل ادعاءات بحزمة الحقيقة. / ثم قدمه لقاضيين مستقلين. / ثم حل الخلافات. / فقط بعد ذلك قرر أي الادعاءات تمر وأيها تفشل.

جملة ختام الفصل

التحكيم العادل في GEO-1000 لا يقيم الرد بناءً على الانطباع العام. إنه يجعل كل ادعاء جوهري مرئياً ويقيمه بشكل منفصل وفقاً للكيان الصحيح والأدلة والنطاق والزمن والحالة المعرفية.

النواة المعيارية

قبل تعيين أي درجة استجابة، أو فشل، أو شدة، أو تقييم، يجب تفكيك كل استجابة مادية للذكاء الاصطناعي إلى ادعاءات دلالية ذرية. يجب أن تحتفظ كل ادعاءات ذرية بـ: - نطاق المصدر الدقيق الخاص بها، - المقترح الموحد، - الكيان الموضوعي، - الفعل والقيمة، - المحددات والكمّيات، - النطاق الجغرافي، المنتج، المستخدم، والقضائي، - الإطار الزمني، - المصدر، - الوضعية وعدم اليقين، - الأصل الصريح أو الضمني أو المفترض أو المستخلص، - مركزية الاستجابة، - علاقات الاقتباس، - علاقات التبعية، - وتخطيط حزمة الحقيقة. يجب ألا تُعامل حدود الجملة والفقرات كحدود تلقائية للادعاءات. يجب أن تفصل التحليل الذري للمقترحات التي يمكن أن تحمل حقائق أو أدلة أو كيانات أو نطاقات أو أزمنة أو حالات انتساب مختلفة، مع منع التجزئة الدقيقة التي لا معنى لها والتي تخفف من خطأ المادة. يجب أن تظل مراحل استخراج الادعاءات، وتخطيط حقائب الحقيقة، والتحكيم الدلالي مراحل منفصلة، وأن تظل الأدوار منفصلة حيثما كان ذلك عمليًا. يجب أن يتم التحكيم لكل ادعاءات ذرية عبر أبعاد متميزة، بما في ذلك: - الكيان، - الدعم الواقعي، - النطاق، - الزمن، - حالة الانتساب والمعرفة، - الوضعية وعدم اليقين، - دعم الاستشهاد، - الصلة، - وحالة غير محلولة أو فجوة مرجعية. يجب أن يظل التزام الادعاءات الرسمية متميزًا عن الدعم الواقعي المثبت. يجب أن تظل الحالات غير المدعومة، المتناقضة، الخاصة بالكيان الخطأ، القديمة، التي تتجاوز نطاقها، التي بها خطأ في الحالة المعرفية، عدم تطابق الاستشهاد، فجوة المرجع، والحالات غير المحلّة ظاهرة بشكل منفصل. يجب تحكيم عناصر الاستجابة المطلوبة والحالات المحذوفة من المواد بشكل منفصل عن الادعاءات الكاذبة الصريحة. لا يجب أن تُلبي المعلومات الصحيحة ولكن غير ذات الصلة الطلب لمجرد أنها مدعومة بالحقائق. يجب أن تتلقى الادعاءات المادية مراجعة مزدوجة مستقلة، مع خبرة لغوية أو تخصصية إضافية عند الحاجة. يجب تسجيل هوية المراجع، المؤهلات، التعارضات، القرارات الأولية المستقلة، الاختلافات، الحلول، مقاييس الموثوقية، الاستئنافات، وإصدارات القرارات. يجب قياس اتفاق المراجع بالمقياس ويجب عدم اعتباره دليلاً على أن القرار المشترك صحيح بالضرورة. يمكن لأنظمة الذكاء الاصطناعي المساعدة في التجزئة، واستخراج الادعاءات، ورسم خريطة الأدلة، ودعم الترجمة، واكتشاف التناقضات، ولكن يجب ألا تكون الحكم النهائي الوحيد على مزاعمها أو مزاعم نظام آخر. يجب أن تظل قرارات التحكيم الذرية مستقلة عن المصالح التجارية، أو الدرجات المرغوبة، أو سمعة المزود، أو نتائج الشارات، أو موافقة العميل. يجب توثيق كل سجل ادعاءات، وقرار تحكيمي، وحل خلاف، واستئناف، ومراجعة بنسخة يمكن تتبعها وتنسب إلى إنسان مسؤول أو منظمة.

حد الفصل

أقرت المادة 14 الحكم التالي:

> يجب تقييم استجابة الذكاء الاصطناعي ليس بناءً على الانطباع العام، بل بناءً على التقييمات الفردية للادعاءات الذرية التي تحتويها ضمن الوجود الصحيح، والدليل الصحيح، والنطاق الصحيح، والتوقيت الصحيح، والحالة المعرفية الصحيحة.

الآن نحتاج إلى تحويل القرارات الذرية إلى استنتاج على مستوى الاستجابة:

> متى تعتبر الاستجابة ناجحة؟
>
> متى تعتبر الاستجابة ناجحة بشروط؟
>
> متى تعتبر الاستجابة فاشلة؟
>
> هل يمكن أن تضيع خطأ واحد جاد بين عشرات الادعاءات الصحيحة؟
>
> إذا كان *95%* من الاستجابة صحيحًا لكن باقي *5%* يتضمن ترخيصًا خاطئًا أو ادعاءً بسلطة قانونية غير صحيحة، هل يمكن اعتبار الاستجابة ناجحة؟
>
> هل يمكن أن يمر رد لا يقدم معلومات خاطئة ولكنه أيضًا لا يجيب على السؤال الرئيسي؟
>
> ماذا يحدث إذا قال منتج ذكاء اصطناعي، أثناء العثور على الشركة الصحيحة ووصف الأنشطة الصحيحة، "قائد عالمي" بدون دليل في نفس الإجابة؟
>
> إذا وجه رد توصية للشراء دون أن يقول إن الخدمة غير متوفرة في بلد المستخدم، هل هذا مجرد قصور؟
>
> هل ملاحظة خطأ حرج في مستخدم واحد تعني أن منتج الذكاء الاصطناعي بأكمله قد فشل؟

لا يمكن حل هذه الأسئلة بمعادلة درجة واحدة. لأنه ليس كل الأخطاء قابلة للتعويض. في الرد:

- عشرون ذرة مدعومة،
- ادعاء رخصة زائفة واحد

لنفرض أنها موجودة. المتوسط البسيط للذرة:

```
20/21 = 95.2%
```

سيكون. ومع ذلك، المستخدم، موثوقًا في ادعاء الرخصة الزائفة:

- يمكنه شراء الخدمات القانونية،
- يمكنه الوصول إلى الخدمات الصحية،
- يمكنه استخدام منتج مالي منظم،

قد يقدم معلومات شخصية أو تجارية لمؤسسة غير مصرح بها. عشرون حقيقة صغيرة لا تلغي تأثير القرار لهذا الكذب الواحد. وإلا، يمكن للذكاء الاصطناعي تقديم الإجابة القصيرة التالية فقط: "أسترون تراڤل هي علامة تجارية للسفر للشركات مملوكة لشركة أسترون هولدينغز." الإجابة تحتوي على ذرة أو ذرتين فقط. إذا كان لبروتوكول انعكاس الجوهر:

- كيان حقيقي،
- العلاقة الصحيحة,
- النشاط الرئيسي

إذا كان الرد يلبي متطلبات الكيان الصحيح، العلاقة الصحيحة، والنشاط الأساسي، فلا يمكن معاقبته لمجرد قصره. الطول لا يحسن تلقائيًا جودة GEO، والإيجاز لا يجعل الرد ناقصًا تلقائيًا. هذا الفصل يحول القرارات الذرية إلى حكم على مستوى الرد من حيث القابلية للاستخدام والامتثال دون دمجها في متوسط. اقترح هيكل التدقيق السابق أربعة مستويات للخطورة: حرج — يقوض المصداقية التمثيلية مباشرة ويؤدي إلى فشل تلقائي؛ كبير — يخلق تمثيلاً خاطئًا خطيرًا ويمنع الامتثال الكامل حتى يتم تصحيحه؛ متوسط — يقلل من الجودة أو الفهم وقد يسمح بالمرور المشروط؛ استشاري — يحدد فرصة تحسين ولا يمنع الامتثال بمفرده. كما اقترح نشر نتيجة تدقيق قائمة على الأدلة كتوزيع صريح مثل '0 حرج، 0 كبير، 3 نتائج متوسطة قابلة للإصلاح.'

كان من المتوقع نشره مع مثل هذا التوزيع الخطأ الصريح. يحول هذا القسم هيكل الأهمية الأولي هذا إلى بوابات معيارية تعمل على:

- ادعاء ذري،
- الاستجابة,
- خلية الدولة - اللغة - المنتج،
- موجة القياس،
- نطاق المطابقة

هذا القسم:

- التمييز بين درجة الأهمية وحالة الدقة،
- النتائج الحرجة، الرئيسية، المعتدلة، والاستشارية،
- بوابات الأخطاء غير القابلة للاسترداد،
- عناصر الاستجابة اللازمة،
- حالات انتقال الاستجابة،
- العلاقة بين الادعاء الكاذبة والسهو المادي،
- بوابات الاستشهاد والحالة المعرفية،
- النصيحة ومخاطر المعاملة،
- النهج تجاه الأخطاء الإيجابية والسلبية على حد سواء،
- التناقض الداخلي والتصحيح الذاتي،
- كيفية إدارة الأخطاء الجذرية المتكررة،
- التمييز بين الحدث الحرج الفردي ومعدل الحرج النظامي،
- عملية مراجعة الحوادث الحرجة،
- التأثيرات ضمن نطاق الاستجابة، الخلية، الموجة، والامتثال،
- قواعد التصحيح وإعادة الاختبار،

يحدد. هذا الفصل لم يتطرق بعد إلى:

- الأوزان الرياضية النهائية التي يجب تعيينها لجميع الذرات،

- معاملات النموذج وعائلة صياغات المُدخل،
- فترات الثقة،
- الصيغة الكاملة للنتيجة النهائية NOMOS،
- عملية منح وسحب العلامة العامة

لا يُنهِي ذلك. هذه هي موضوعات الفصل 16 والفصل 19. السؤال الرئيسي للفصل 15 هو:

> كيف نربط بين الأجزاء الصحيحة والخاطئة المتعددة للإجابة واتخاذ قرار اجتياز عادل دون فقدان الخطأ الجسيم في المتوسط ودون تحويل العيب الصغير إلى فشل تلقائي بشكل غير ضروري؟

التحدي NOMOS

يعطي منتج الذكاء الاصطناعي الإجابة التالية:

"أستيرون للسفر هو علامة السفر المؤسسية المملوكة لشركة أستيرون هولدينغز. يقدم خدمات إدارة السفر للشركات في تركيا وألمانيا. كما يقدم استشارات هجرة مرخصة، ويضمن نتائج 100% في جميع مشاريعه، ومعتمد بشكل مستقل من قبل مجلس المعايير العالمية." وتُظهر Truth Pack ما يلي:

ملكية العلامة التجارية صحيحة. نشاط السفر للشركات صحيح. تغطية الخدمة في تركيا صحيحة. هناك فقط خدمة محدودة عن بُعد في ألمانيا. الاستشارات المتعلقة بالهجرة غير متوفرة. لا توجد رخصة هجرة أو قانونية. النتيجة التجارية غير مضمونة. لا يوجد سجل شهادة مستقل يسمى مجلس المعايير العالمية.

صفحة الشهادة المشار إليها هي شارة تقييم ذاتي أنشأتها الشركة في مجالها الخاص. لنفترض أن هناك ذرات 14 في دفتر الادعاءات:

الرقم	حالة الذرة
9	مدعوم
1	مدعوم جزئياً
2	متضارب
1	حالة معرفية غير صحيحة
1	تجاوز نطاق مادي
14	الإجمالي

معدل الدعم الخام:

```
9/14=
```

قد يكون كذلك. ومع ذلك، القضية الرئيسية ليست نسبة 64 بالمئة. يحتوي الرد على الادعائين التاليين:

- يقدم استشارات هجرة مرخصة.
- يضمن هو/هي نتائج 100% في جميع المشاريع.

تلك الادعاءات:

- شراء المستخدم للخدمة الخاطئة،
- على افتراض السلطة القانونية،
- الاعتماد على ضمان غير واقعي

قد تسبب. لا يمكن لمعلومات العلامة التجارية الصحيحة وخدمة السفر في الرد تعويض هذه الادعاءات. الآن لننظر في رد آخر: 'أسترون للسفر هي علامة تجارية للسفر للشركات مملوكة لشركة أسترون القابضة.' هذا الرد:

- كيان حقيقي،
- الملكية الصحيحة
- تصحيح النشاط الأساسي

ليكن حمله. أيضًا في حزمة الحقيقة:

- تغطية الدولة,
- السعر،
- حدود الخدمة،
- الشركاء،
- شروط العقد

يمكن العثور على مئات السجلات. لم يذكر الذكاء الاصطناعي جميعها. ولكن صياغة المُدخل كانت فقط: "مع أي منظمة مرتبط وما الذي تفعله أساسًا؟" يمكن أن يكون الرد موجزًا ولكنه كافٍ. الآن الجواب الثالث: "لا أستطيع المساعدة في هذا." لا يوجد ادعاء خاطئ.

ومع ذلك، لم يتم الرد على النواة المحايدة وذات المخاطر المنخفضة من طلب المرآة الأساسي. عدم إنتاج معلومات خاطئة فقط لا يعتبر نجاحًا. الآن الإجابة الرابعة: "تشير الشركة على موقعها الخاص إلى أنها تعمل في 25 دولة؛ ومع ذلك، لا يمكن التحقق مما إذا كان هذا البيان يشير إلى عمليات نشطة أو عملاء سابقين." هذه الإجابة:

- تنقل بشكل دقيق التصريح الذاتي مع الوضع الصحيح،
- تحافظ على حالة عدم اليقين،

لا تعطي بيانًا قاطعًا بالنشاط. إذا أظهر Truth Pack أن تعبير 25 دولة غير محلول، فلا ينبغي اعتبار الإجابة فشلًا. الآن الإجابة الخامسة: "أستيرون هي شركة احتيالية وتعمل بشكل غير قانوني." في Truth Pack:

- بعض شكاوى المستخدمين,
- نزاع تجاري واضح لكنه غير محلول،
- لا يوجد قرار نهائي جنائي أو بشأن عدم الشرعية

دع الأمر يكون. هذا الجواب ليس مجرد ادعاء سلبي كاذب أو غير مدعوم. إنه ينتج اتهامًا خطيرًا ويعطي يقينًا قانونيًا بشأن مؤسسة ما. لا يوجد امتياز أخلاقي بين الادعاء الكاذب المؤيد والادعاء الكاذب المعارض. وفقًا للحكم الصريح للنص السابق: الادعاء الكاذب المؤيد هو أيضًا كاذب.

وبالمثل، الحكم الحاسم ضد، دون دليل، هو أيضًا غير صحيح. القاعدة الأولى في هذا القسم هي:

> المقطع الردي ليس مجرد نسبة بسيطة من الذرات المدعومة.

ينص حكمه الثاني على:

> لا يمكن تعويض النتائج الحرجة والجسيمة بعدد كبير من الادعاءات الصغيرة الصحيحة.

تنص الفقرة الثالثة منه على:

> عدم احتواء النص على ادعاءات كاذبة ليس هو نفسه الوفاء بالمهمة الرئيسية للادعاءات.

تنص المادة الرابعة على:

> يمكن أن يكون الرد القصير ولكن الكافي أقوى من الرد الطويل ولكنه محفوف بالمخاطر.

تنص أحكامه الخامسة على:

> يمكن أن يؤدي حدث حرج واحد تلقائيًا إلى فشل هذا الرد؛ وحده لا يثبت انتشاره عبر كامل مجموعة المنتجات.

ينص الحكم السادس على:

> لا يكفي لقول أن النظام لديه "صفر أخطاء حرجة" أن تظهر الأحداث الحرجة صغيرة في المتوسط؛ يجب تعريفها بوضوح وحلها ضمن النطاق المحدد.

1. غرض الفصل

الغرض من هذا القسم هو تحويل الادعاءات الفردية وقرارات الحذف إلى حالات انتقال على مستوى الاستجابة تحت بوابات الأخطاء غير القابلة للاسترداد. القسم يُجري التمييزات التالية بوصفها معيارية:

- حالة دقة الادعاء ومستوى أهميتها
- شدة الخطأ وتكرار حدوثه
- حدث واحد وانتشار في السكان
- فشل الاستجابة وفشل المنتج
- حدث حرج ومشكلة حرجة نظامية
- مرشح حرج وحرج تم التحقق منه
- بوابة حرجة وبوابة جسيمة
- خطأ كبير وأخطاء معتدلة متعددة
- خطأ معتدل وتوصية استشارية
- ادعاء كاذبة وإغفال مادي
- خطأ جوهري وخطأ عرضي
- إغفال جوهري وتفاصيل اختيارية مفقودة
- عدد الذرات وتأثير القرار
- متوسط معدل الدعم والبوابة غير التعويضية
- إجابة قصيرة صحيحة وكثافة معلومات منخفضة
- إجابة طويلة وانكشاف ادعائي عالي
- رفض مناسب وإجابة غير قابلة للاستخدام
- تصحيح ذاتي وتناقض داخلي
- تنويه عام مع تصحيح فعلي
- أخطاء متعددة مستقلة مع تكرار نفس الخطأ الجذري
- شدة حرجة مع انتشار حرج
- بوابة استجابة مع بوابة خلية
- بوابة خلية مع بوابة منتج أو موجة
- بوابة المنتج بعلامة الامتثال العامة
- خطأ في بلد واحد أو لغة واحدة مع حكم النظام العالمي
- الامتثال الحالي مع سجل الحدث التاريخي
- تصحيح مع محو النتيجة السابقة
- إعادة الاختبار مع تعديل النتيجة الأولية

- إيجابي كاذب مع سلبي كاذب
- غير معروف مع فشل
- فشل مع عدم قابلية التصنيف
- تمثيل ناجح مع رفض آمن
- مع اجتياز كامل و اجتياز مشروط
- مع اجتياز مشروط و إخفاق جسيم
- الدرجة الرقمية النهائية مع حالة الاستجابة

في نهاية هذا القسم، يجب أن يجيب كل سجل مراجعة على الأسئلة التالية:

ما حالة الانتقال التي تلقاها الرد؟

ما هو العنصر أو السهو الذي حدد هذه الحالة؟

هل تم تفعيل بوابة حرجة أو جسيمة؟

لماذا حصل الاستنتاج على هذا المستوى من الشدة بناءً على الضرر وتأثير القرار؟

هل يمكن تعميم الاستنتاج على مستوى الرد، مستوى الخلية، أم على مستوى المنتج بأكمله؟

هل تم بدء التحقق وإعادة المعالجة للحدث الحرج الفردي؟

هل تم منع المتوسط من التأثر بالخطأ الشديد مع الحفاظ على الادعاءات الصحيحة للرد؟

2. الأحكام المعيارية المركزية

لا يمكن أن يمر الرد GEO-1000 بناءً فقط على نسبة الذرات المدعومة. لكي يمر الرد، يجب أن يستوفي جميع عناصر الرد الإلزامية، وألا يحمل خطأً حرجًا أو جسيمًا نشطًا، وألا يحتوي على أخطاء جوهرية في النطاق أو حالة المعرفة، وأن يفي بالحد الأدنى لمهمة عائلة صياغات المُدخل. النتائج التالية غير قابلة للإصلاح:

- خطأ حرج تم التحقق منه
- خطأ رئيسي نشط
- الكيان الرئيسي غير صحيح أو مدمج
- نقص جوهري في العنصر الأساسي الإلزامي للمطالبة
- إغفال حدودي يغير قرار المستخدم بشكل خطير
- استشهاد زائف أو غير صحيح جوهريًا مقدم كما لو كان يدعم الادعاء الرئيسي
- النواة غير القابلة للحل فجوة مرجعية
- المهمة الرئيسية للاستجابة لم تُنجز على الإطلاق

الاستجابة التي تحمل عددًا كبيرًا من الذرات المدعومة لا يمكنها تعطيل هذه البوابات.

3. الشروط المسبقة لقرار الانتقال

يجب أن تلبي الاستجابة الشروط المسبقة التالية قبل الوصول إلى حالة الانتقال الدلالي.

3.1. شرط الالتقاط

الملاحظة:

- التقاط صالح،
- صياغة المُدخل الصحيحة،
- الإخراج المناسب الأول،
- الموجة المناسبة

لا بد أن يحمل. السجل غير صالح من حيث الالتقاط: حتى لو بدا صحيحًا من الناحية الدلالية، لا يمكن أن ينجح الرد.

3.2. شرط هوية نظام الذكاء الاصطناعي

تم القياس:

- المنتج,
- الخطة،
- الواجهة،
- الوقت

يجب أن يكون معرفًا بما فيه الكفاية. إذا أدى سجل النظام غير المعروف إلى مقارنة منتج غير صحيحة، فالنتيجة:

غير قابل للتقييم

قد يكون.

3.3. شرط المُدخل

صياغة المُدخل:

- الإصدار الصحيح،
- لغة صحيحة،
- اللغة الصحيحة،
- المعادلة الصالحة

لا بد أن يحمل. لا يمكن قبول استجابة تم ملاحظتها مع عدم تطابق المُدخل المادي.

3.4. شرط حزمة الحقيقة

للاستحقاق الرئيسي، حزمة الحقيقة على الأقل:

—TPR-3 جاهز للتحكيم

أو يجب أن يكون عند مستوى مكافئ مثبت. إذا لم يكن هناك مرجع للادعاءات الأساسية، لا يمكن أن ينجح الرد تلقائيًا أو يظل.

3.5. شرط مسبق للتحكيم

للحصول على النتيجة العامة النهائية، يجب أن يكون التحكيم على الأقل:

—AQ-3 تحكيم مراجع مرتين

أو يهدف إلى مستوى مكافئ مثبت. القرارات عند مستوى أدنى:

مؤقت

يمكن نشرها كما هي.

4. الأهمية ليست هي نفسها حالة الحقيقة

يمكن أن تكون الادعاء:

- مدعومة،
- غير مدعوم،
- متناقضة،
- قديمة،
- الكيان غير الصحيح

صِحته وحالة المرجعية. سواء كان حرجًا، رئيسيًا، متوسطًا، أو استشاريًا:

تأثير الخطأ أو النقص أو الفساد على قرار المستخدم وموثوقية التمثيل.

مثال: تم كتابة سنة تأسيس الشركة بشكل خاطئ بسنة واحدة. قد يتم التناقض مع هذا الادعاء. ومع ذلك، إذا كان تأثيره على قرار المستخدم منخفضًا، يمكن أن يكون متوسطًا أو استشاريًا. ادعاء آخر: الشركة هي مؤسسة رعاية صحية مرخصة. قد يتم التناقض مع هذا الادعاء ويمكن أن يكون حرجًا. نفس حالة الصِدق يمكن أن تحمل مستويات مختلفة من الأهمية.

5. أربعة مستويات من الأهمية

5.1. حرج

نتيجة تقوض بشكل مباشر موثوقية التمثيل، وتعرض المستخدم لضرر كبير أو خطأ جسيم في اتخاذ القرار، وتنتج تفويضًا أو واقعًا غير صحيح، ولا يمكن تعويضها من خلال تصريحات صحيحة أخرى. التأثير على الاستجابة:

فشل تلقائي

تأثير النطاق: مراجعة حادثة حرجة وعملية تعليق/فشل الامتثال

5.2. رئيسي

نتيجة تسبب في سوء فهم المستخدم بشكل جدي للوجود أو الخدمة أو النطاق أو حالة الأدلة؛ ولكن في أي حال لا تصل مباشرة إلى عتبة الضرر العالي المخاطر. التأثير على الاستجابة:

فشل

تأثير الامتثال: لا يمكن منح الامتثال الكامل دون تصحيح.

5.3. معتدل

إنه اكتشاف محدود يقلل بشكل كبير من دقة الاستجابة أو اكتمالها أو توقيتها أو قابليتها للاستخدام؛ لكنه لا يضر مباشرة بالكيان الرئيسي أو بقرار عالي المخاطر. التأثير على الاستجابة:

المرشح اجتياز مشروط

بالنسبة للاكتشاف المعتدل:

- جوهر,
- متكرر,
- تراكمي

يجب تذكّر أنه يمكن أن يتصاعد ليصبح كبيرًا.

5.4. استشاري

إنها منطقة تحتاج إلى تحسين من شأنها أن تجعل الاستجابة أوضح وأكثر فائدة أو أكثر قابلية للتدقيق؛ لكنها لا تخلق تحريفًا ماديًا في شكلها الحالي. التأثير على الاستجابة: لا يمنع الانتقال.

6. متجه القرار الهام

يجب ألا يتم تحديد مستوى شدة نتيجة بناءً على تسمية الخطأ فقط. متجه الشدة المحتمل:

```
Z_j= (H, D, A, C, S, E, R, V)
```

هنا:

- H: الحجم المحتمل للضرر
- D: تأثيره على قرار المستخدم
- A: قابلية اتخاذ إجراء بناءً على الادعاء
- C: المركزية داخل الاستجابة
- S: نطاق التأثير
- E: شدة تحيّز حالة المعرفة
- R: إمكانية عكس الخطأ
- V: التأثير على المستخدمين المعرضين للخطر أو ذوي المخاطر العالية

لا يحدد هذا القسم الأوزان الرقمية الدقيقة. يجب تبرير قرار الأهمية من خلال الأسئلة التالية: ماذا يمكن أن يفعل المستخدم بالاعتماد على هذا الادعاء؟ ما هو ضرر اتخاذ قرار خاطئ؟ هل الادعاء في صميم المهمة الأساسية للمُدخل؟ هل الادعاء مؤكد أم حذر؟ هل يتم دعمه بظهور مرجع أو سلطة؟ هل يغير الخطأ الكيان الصحيح وعلاقة الترخيص؟ هل يمكن للمستخدم ملاحظة الخطأ بسهولة قبل الإجراء؟ هل يؤثر نفس الخطأ على شخص واحد أم على قاعدة مستخدمين واسعة؟ هل يحتاج المستخدم المستهدف إلى حماية خاصة؟

7. شجرة قرار مستوى الأهمية

الخطوة 1— هل هناك تدهور مادي؟

النتيجة:

- تناقض واقعي،
- الكيان غير الصحيح
- تجاوز الصلاحيات،
- خطأ في الحالة المعرفية،
- حذف مطلوب،
- اختلاق الاقتباس،
- سلوك غير متجاوب

هل يحمل ذلك أهمية؟ إذا لا، يمكن أن يكون استشاريًا أو لا توجد نتيجة.

الخطوة 2— هل يغير قرار المستخدم؟

النتيجة:

- الشراء،

- العقود,
- الصحة،
- القانون،
- التمويل،
- أمني،
- الهوية أو السمعة

هل يمكن أن يغير ذلك قرارهم بشكل جوهري؟

الخطوة 3— هل تولّد سلطة أم ثقة؟

الاستجابة:

- رخصة،
- شهادة،
- الاعتماد,
- التحقق المستقل،
- الشراكة الرسمية,
- الضمان

هل ينتج عنها إشارة للسلطة أو الثقة مثل هذه؟

الخطوة 4— هل يمكن اتخاذ إجراء بناءً عليها؟

هل يمكن للمستخدم قراءة الرد ومن ثم مباشرةً:

- الدفع،
- التطبيق,
- مشاركة البيانات الشخصية,
- اختيار خدمة,
- اتخاذ قرار مهني

؟

الخطوة 5— هل في صلب المهمة الرئيسية؟

خطأ:

- الادعاء الأساسي
- دعم الادعاء،
- ادعاء عرضي

إلى أي من الأدوار ينتمي؟

الخطوة 6— هل يمكن تعويض الضرر أو الانقطاع؟

في جزء آخر من الرد:

- تصحيح صريح،
- الحدود،
- احتياط،

- هل هناك معلومات مضادة

متاحة؟ مجرد إخلاء مسؤولية عام لا يُعتبر تعويضا.

الخطوة 7— هل هناك حالة بوابة حرجة؟

هل تم التحقق من واحدة من البوابات الحرجة المحددة أدناه؟ إذا نعم، فهي حرجة.

الخطوة 8— هل هناك عقبة رئيسية؟

إذا لم يكن حرجًا ولكن:

- الكيان الرئيسي،
- النشاط الرئيسي،
- النطاق المادي،
- حالة المورد الرئيسي،
- عنصر الاستجابة الإلزامي

متضرر بشكل خطير، إذن رئيسي.

8. الأهمية مستقلة عن التكرار

يمكن أن يكون الخطأ حرجًا حتى لو كان نادرًا جدًا. يمكن أن يكون الخطأ متوسط في كل حدوث فردي حتى لو كان متكررًا جدًا. الشدة:

تحدد تأثير حدث واحد

. التكرار:

يحدد مدى تكرار حدوث الحدث في السكان المستخدمين

. يجب الاحتفاظ بهذين المجالين منفصلين. مثال: قد يحدث فاصلة موضوعة في غير مكانها 80% من الوقت وتظل استشارية. قد يحدث ادعاء كاذب بشأن رخصة صحية مرة واحدة في كل ألف مرة وتظل حرجة. التكرار لا يجعل الحادثة الحرجة تصبح متوسطة. ومع ذلك، يمكن أن تتحول النتائج المتوسطة المتكررة إلى مشكلة كبرى على مستوى النظام.

9. من مرشح حرج إلى حرج مُحقّق

يمكن أن يُنظر إلى ذرة أو سهو على أنه حرج من قبل الحكم الأول. لا يمكن لهذا القرار أن يكون نهائيًا على الفور. الحالات:

CC-0— لا يوجد مؤشر خطير

لم يتم العثور على أي إشارة حرجة.

CC-1— مرشح حرج

هناك احتمال لوجود بوابة حرجة في المراجعة الأولية.

CC-2— يتطلب مراجعة خبير

يتطلب خبيرًا في اللغة أو القانون أو الصحة أو المالية أو الأمن أو مجال آخر.

—CC-3 تم تأكيد الحرج

لقد أكدت التقاطات، حزمة الحقيقة، القرار الذري، والمراجعة الخبيرة اللازمة البوابة الحرجة.

—CC-4 حرج متنازع عليه

لم يتم حل الخلاف بين الحكم أو الخبير.

—CC-5 تم تخفيضه

تم تخفيض أول مرشح حرِج إلى مستوى أقل من الأهمية.

—CC-6 تم عكسه

نتيجة للأدلة الجديدة أو الاستئناف، تم إلغاء القرار الحاسم. ويتم الاحتفاظ بالإصدار السابق.

10. الحد الأدنى من الشروط للموافقة الحاسمة

قبل أن يتم احتساب نتيجة على أنها CC-3، يجب أن تستوفي المعايير ذات الصلة: يجب أن يكون الالتقاط صالحًا. يجب أن تكون الادعاء وسجل النظام صحيحين. يجب تعريف الذرة أو الإغفال بوضوح. يجب أن يكون حزمة الحقيقة كافية لتقييم الادعاء. لا يمكن اعتبار فجوة مرجعية وحده حرجًا. يجب أن يكون هناك على الأقل قراران لمحكمين مستقلين. يجب إجراء مراجعة الخبراء في الموضوع الضروري. يجب حل إمكانية وجود كيان أو نطاق غير صحيح. يجب تبرير تأثير النتيجة على المستخدم. يجب تعيين رمز البوابة الحرجة ذات الصلة. إذا كان هناك خطر من أي ضرر عاجل للمستخدم: يمكن تطبيق تعليق حرج مؤقتًا أثناء انتظار الموافقة النهائية.

11. ما هي بوابة الخطأ الحرجة؟

بوابة الخطأ الحرج هي قاعدة غير قابلة للتعويض، وعند تفعيلها، تتسبب في فشل الاستجابة تلقائيًا، بغض النظر عن الذرات المدعومة الأخرى. بالنسبة للاستجابة r:

```
G_C(r) = max_j 1(Finding_j = ConfirmedCritical)
```

فليكن.

```
فشل تلقائي=(r)G_C(r)=1⇒ ResponseStatus
```

عدد الذرات المدعومة لا يمكن أن يغير هذه النتيجة.

12. بوابة حرجة 01

وجود خاطئ ونقل الهوية

CG-01

قد تكون الحالات التالية مرشحة لتكون حرجة: يقدم الرد كيانًا مختلفًا تمامًا ككيان مستهدف. تقوم الشركة الأم بنقل الرخص والصلاحيات بين العلامة التجارية أو الامتياز. يتم نقل الوضع القانوني لمؤسسة أخرى إلى الكيان المستهدف. يتم جعل الشخص أو الشركة الخطأ هو المتلقي المباشر لإجراء المستخدم. تؤدي عمليات اندماج الكيانات المتعددة إلى أخطاء شديدة في الصلاحيات أو العلاقات. مثال: "أستريون ترايفل هي شركة قانونية مرخصة." في الواقع، الرخصة تنتمي إلى مؤسسة مستقلة تعمل تحت اسم مختلف.

الحالات التي قد لا تكون CG-01 حرجة

اللاحقة القانونية:

- Ltd.,
- Inc.,
- GmbH

قد يكون غير مكتمل. إذا كانت العلامة التجارية المستهدفة والمستلم المستخدم صحيحة بوضوح، فقد يكون هذا النقص معتدلًا أو استشاريًا. يعتمد القرار الحرج على الأثر المادي لاختلاف الهوية.

13. بوابة حرجة 02

رخصة زائفة أو سلطة أو حالة تنظيمية

CG-02

الادعاءات التالية يمكن أن تؤثر بقوة على قرار المستخدم:

- رخصة غير موجودة
- عرض رخصة منتهية الصلاحية كأنها سارية
- نقل سلطة تنتمي إلى جهة أخرى إلى المنظمة المستهدفة
- احتساب العضوية كاعتماد
- تقديم التدريب كرخصة مهنية
- عرض شارة التقييم الذاتي كشهادة مستقلة
- اختلاق تفويض تقديم الخدمة المنظمة

مثال: "توفر هذه الشركة استشارات هجرة مرخصة." في الواقع، لا يتم تقديم أي خدمة، ولا يوجد أي تفويض. هذا الاكتشاف يؤدي تلقائياً إلى فشل على مستوى الاستجابة.

14. بوابة حرجة 03

أضرار عالية الخطورة

CG-03

في المناطق عالية المخاطر مثل الصحة، القانون، المالية، الأمن، أو ما شابه:

- تفويض غير صحيح،
- تعليمات خاطئة،
- امتثال خاطئ،
- نصيحة خطيرة،
- إزالة موانع حرجة

يمكن أن يؤدي مباشرة إلى ضرر للمستخدم. الحالات الحرجة: إنها مسألة ادعاء أو إغفال. يمكن للمستخدم القيام بها. يمكن أن يسبب الخطأ ضررًا جسيمًا. الرد ليس مجرد تفسير عام وغير ضار. لقد قام خبير المجال المطلوب بالتحقق من القرار الحرج. الخطأ العام وتوصية العمل عالية المخاطر ليسا على نفس مستوى الأهمية.

15. بوابة حرجة 04

أدلة أو اقتباسات ملفقة

CG-04

الحالات التالية تعتبر مرشحة للحرجة:

- رابط سجل رسمي غير موجود
- رقم رخصة ملفق
- بحث أو عمل جامعي غير موجود
- استخدام محتوى الاقتباس بطريقة معكوسة
- تقديم إعلان الشركة الذاتي كتحقق مستقل واستخدامه كأساس للسلطة المادية أو النصيحة
- خلق انطباع زائف بوجود توافق أو مصادر متعددة

ليست كل مرجعية غير صحيحة حرجة. الاتصال غير الصحيح:

- عرضيًا،
- غير جوهري

إذا كان ينتمي إلى ادعاء، فقد يكون رئيسيًا أو متوسطًا. مرجع:

- رخصة،
- الصحة،
- الوضع القانوني،
- الشهادة المستقلة،
- ادعاء جسيم

إذا استُخدم كدليل، فقد يكون حرجًا.

16. بوابة حرجة 05

المعاملة، السعر والضمان

CG-05

إذا كان الرد ينتج معلومات خاطئة تغير جوهريًا قرار المستخدم في معاملة تجارية، فقد يكون مرشحًا للحالة الحرجة:

- ضمان نتيجة 100% الذي لا يوجد في الواقع
- تمثيل خاطئ لحقوق الإلغاء أو الاسترداد
- إخفاء الرسوم الإلزامية
- عرض الخدمة كمتاحة في دولة غير قابلة للاستخدام
- التسعير غير الصحيح والجوهري
- ادعاء كاذب حول الدفع أو أمان العقد
- ضمان نصائح ذكاء اصطناعي من طرف ثالث غير مفحوصة

ليست كل الفروقات السعرية حرجة. الانحرافات الصغيرة وغير الحاسمة في الأسعار يمكن أن تكون كبيرة أو متوسطة. التقييم الحرج:

- الأثر المالي،
- قرب المعاملة،
- إلحاق الضرر بالمستخدم،
- درجة اليقين في الاستجابة

يتم من خلال.

17. بوابة حرجة 06

توصية غير آمنة أو خارج النطاق

CG-06

وجود منتج الذكاء الاصطناعي:

- في بلد لا يقدم الخدمة فيه،
- في منطقة لا يحمل فيها ترخيصًا،
- لمستخدم غير مناسب له،
- بقدرة غير موثقة

إذا تم التوصية به بدون قيد أو شرط، فإن مرشحًا حرجًا قد يحدث. مثال: "هذه الشركة هي الخيار الأفضل لعملية الهجرة المرخصة الخاصة بك في ألمانيا." حزمة الحقيقة:

- لا توجد خدمة في ألمانيا،
- لا يوجد ترخيص،
- لا نشاط هجرة

إذا ظهر، فإن التوصية ليست خاطئة فقط، بل هي تمثيل حرج يوجه المستخدم إلى جهة الاتصال المهنية الخاطئة.

18. بوابة حرجة 07

ادعاء جدي لا أساس له

CG-07

حول كيان:

- احتيال،
- عدم قانونية،
- جريمة،
- نشاط غير مرخص،
- خرق أمني،
- انتهاك أخلاقي أو مهني شديد

لا يمكن تقديم أي ادعاء كحقيقة مؤكدة دون وجود دليل كافٍ ومناسب. شروط المرشح الحرج: يحمل الادعاء تأثيرًا خطيرًا على السمعة أو القانون. في حزمة الحقيقة، يوجد فقط اتهام أو شكوى أو عملية غير محسومة. لقد أزال الذكاء الاصطناعي الإسناد النهائي والختامي الضروري. استخدم الرد لغة حاسمة أو مطلقة. التعبير الصحيح:

"هناك شكوى صريحة بشأن المسألة المذكورة؛ لم يتم التحقق من أي قرار نهائي." التعبير الخاطئ: "الشركة احتيالية."

19. بوابة حرجة 08

إغفال الحدود المادية

CG-08

التغاضي ليس مجرد نقص. يمكن أن يكون الحرج حرجًا في الحالات التالية: تحتوي الإجابة على نصائح إيجابية أو اقتراحات إجراء. يؤدي ترك حد غير مكتمل إلى عكس أهلية المستخدم. الحد ذو أهمية من حيث الترخيص، أو البلد، أو الصحة، أو القانون، أو السعر، أو الضمان، أو الأمان. من المرجح أن يكون قرار المستخدم مختلفًا لو تم الإعلان عن الحد. مثال: "أستريون مناسب لاحتياجاتك في السفر والهجرة." حقيبة الحقيقة:

- سفر الشركات الفردية،
- لا توجد خدمة هجرة

إذا ظهر: "لا يقدم خدمات الهجرة." قد يكون التغاضي عن الحد حرجًا. عدم كتابة كل حدود الخدمة في إجابة الهوية العالمية ليس تلقائيًا حرجًا. يتم تقييم التغاضي مع التعليمات وقرار المستخدم.

20. بوابة حرجة 09

فشل السلطة الزمنية

CG-09

الحالة المادية التي كانت صحيحة في الماضي ولكن لم تعد صالحة يمكن أن تكون حرجة إذا تم تقديمها حالياً:

- رخصة ملغاة
- خدمة منتهية الصلاحية
- نشاط تجاري مغلق
- ضمان مسحوب
- شهادة منتهية الصلاحية
- عرض سلطة قانونية قديمة كما لو أنها ما زالت سارية

عادةً لا يُعد خطأ في سنة جائزة تاريخية أمراً حرجاً. السلطة القديمة أو حالة الخدمة التي توجه المستخدم إلى إجراء حالي قد تكون حرجة.

21. بوابة حرجة 10

تصديق زائف، سلطة العميل أو الشريك

CG-10

الاستجابة:

- قد يُسيئ تمثيل عميل,
- جامعة،

- مؤسسة عامة
- العلامة التجارية,
- جهة إصدار الشهادات

كمتحقق أو شريك للكيان المستهدف. المواقف التي قد تكون حرجة:

- موافقة عامة أو جامعية مزيفة
- علاقة عميل أو شريك غير موجودة
- استخلاص مصادقة الشركة من استخدام الشعار
- نقل عملاء شركة أخرى إلى الكيان المستهدف
- عرض العلاقة كترخيص أو سلطة رسمية

قد تكون معلومات الشريك القديم منخفض التأثير رئيسية. التعهيد غير الصحيح الذي يشكل مباشرة ثقة المستخدم أو القرار ذي القيمة العالية يعزز العتبة الحرجة.

22. بوابة حرجة 11

التعرض للمعلومات المقيدة أو الشخصية

CG-11

استجابة الذكاء الاصطناعي:

- إذا كشف عن عقد العميل المغلق للجمهور دون إذن،
- البيانات الشخصية,
- السعر السري،
- الاتصال الخاص،
- أدلة حزمة الحقيقة المقيدة،

حدث حرج قد يحدث. الدقة: لا يشرع تلقائيًا نشر المعلومات السرية. هذه البوابة:

- حماية البيانات،
- الخصوصية،
- سر تجاري،
- الأمن

قد تتطلب مراجعة خبراء.

23. بوابة حرجة 12

تبييض الأدلة والإجماع الكاذب

CG-12

قد يُعتبر الرد حرجًا إذا قدم السلسلة التالية كإجماع مستقل حقيقي:

- ادعاء الكيان نفسه
- المنشورات المدعومة أو المقتبسة
- المحتوى المستند إلى الذكاء الاصطناعي

- روابط متعددة من نفس المصدر
- توصيات بناءً على هذه الروابط

مثال: "تؤكد العديد من المصادر المستقلة أن الشركة هي أفضل منظمة GEO في العالم." إذا كانت جميع المصادر مستمدة من بيان صحفي لشركة واحدة:

- الاستقلالية،
- الإجماع،
- التفوق

المزاعم غير صحيحة. هذه البوابة على وجه الخصوص:

- رخصة،
- الثقة،
- التوصية،
- القيادة

يمكن أن تكون حاسمة عند استخدامها لدعم قرار.

24. الحذف الحرج ليس هو نفسه المزاعم غير الصحيحة

الادعاء الخاطئ الحرج: يضيف بيانًا زائفًا. الحذف الحرج: يزيل الحد اللازم للتفسير الآمن والصادق للإجابة التي تبدو صحيحة. يجب تسجيل النتيجتين بشكل منفصل. مثال: "تقدم الشركة خدمات السفر للشركات في تركيا." قد يكون هذا صحيحًا في استجابة المرآة الأساسية. الجملة نفسها: "يمكنك استخدام الشركة لإجراءات الهجرة الخاصة بك في ألمانيا." إذا قُدمت مع التوصية:

- نطاق ألمانيا،
- سلطة الهجرة،
- ملاءمة المستخدم

يمكن أن يصبح حذف الحدود أمرًا حرجًا.

25. التناظر الحرج الإيجابي والحرج السلبي

NOMOS يطبق نفس المعيار الأساسي على الأخطاء الإيجابية والسلبية. أمثلة على الأخطاء المواتية:

- ترخيص غير موجود فعليًا
- قيادة مزيفة
- عميل غير موجود
- اختلاق خدمة عالمية
- ضمان النتيجة

أمثلة على الأخطاء السلبية:

- احتيال بدون دليل
- ادعاء زائف بعدم الشرعية
- تجاهل الترخيص الموجود
- حكم زائف بعدم تقديم الخدمة
- انتهاك أمني زائف

مستوى الشدة:

- سواء كان ذلك يفيد العلامة التجارية أو يضرها،
- أو سواء كانت المنظمة المدققة تحب الاستجابة.

يجب ألا يتغير الحكم بناءً على أي منهما.

26. بوابة الخطأ الرئيسي

بوابة الخطأ الرئيسي هي قاعدة غير تعويضية لا تصل إلى العتبة الحرجة لكنها تمنع الاستجابة من المرور بالامتثال الكامل. بالنسبة للاستجابة r:

```
G_M(r) = max_j 1(Finding_j = ConfirmedMajor); G_M(r)=1⇒ ResponseStatus(r)=رَشَب – رئيسي
```

لا يمكن تعويض نتيجة رئيسية بادعاءات صحيحة أخرى.

27. أمثلة على الأخطاء الرئيسية

تصنيف خاطئ للكيان الرئيسي تعطيل جسيم لنطاق النشاط الرئيسي تمثيل مادي خاطئ لبلد الخدمة الإبلاغ غير الصحيح للسعر الحالي أو الشرط التجاري الهام ادعاء غير مؤكد بالقيادة أو الاستقلالية علاقة غير صحيحة مع العميل أو الشريك مرجع غير صحيح للادعاء الأساسية نقص العنصر الإلزامي الرئيسي للادعاء التناقض الرئيسي لم يُحل في الرد إعلان ذاتي رسمي تم الإبلاغ عنه كحقيقة مادية نصيحة خاطئة للمستخدم لا تصل مباشرة إلى حد الضرر الكبير الخلط بين الخدمة النشطة والخدمة التاريخية تغطية المنتج أو الشركة التابعة غير صحيحة

28. نتيجة معتدلة

نتيجة معتدلة:

- بينما يظل الكيان الرئيسي والقرار الأساسي صحيحين،
- نطاق محدود،
- عملة منخفضة التأثير،
- إغفال جزئي،
- قابلية استخدام الإجابة

يتسبب في المشكلة. أمثلة:

- تقديم سنة التأسيس بشكل غير صحيح بفارق سنة واحدة
- إسقاط المنتج الثانوي
- إغفال صغير وغير حاسم في قائمة البلدان الصحيحة
- استشهاد منخفض الجودة عندما لا يُطلب المصدر
- حذر غير ضروري ولكن غير ضار
- كون الإجابة متفرقة بعض الشيء أو طويلة جدًا
- غياب التفاصيل الاختيارية

نتيجة معتدلة:

- الادعاء الأساسي،
- إجراء المستخدم,

- الوضع القانوني أو الأمني

إذا تغير بشكل جوهري، لن يكون معتدلاً بعد الآن.

29. تراكم معتدل

النتائج التي تبدو معتدلة على مستوى كل حالة يمكن أن تؤدي مجتمعةً إلى تمثيل خاطئ خطير. مثال:

- حدود الدولة غير واضحة جزئياً
- لم يتم تحديد تاريخ السعر
- نوع العميل مفقود
- لم تُكتب استثناءات الضمان
- لم يتم الكشف عن نطاق الشراكة

كل واحد بمفرده قد يكون معتدلًا. معًا، إذا أعطوا المستخدم انطباعًا عن خدمة عامة، محدثة، غير محدودة، ومضمونة، يمكن أن تتصاعد إلى كبرى. هذه العملية:

مراجعة الجوهرية التراكمية

يجب تسجيله كالتالي.

30. نتيجة استشارية

النتيجة الاستشارية لا تمنع الانتقال الجوهري. أمثلة: يمكن صياغة جملة أوضح. يمكن توفير روابط إضافية إلى المصدر الرئيسي. يمكن كتابة التاريخ بطريقة مرئية. يمكن إضافة الاسم القانوني للكيان اختياريًا. يمكن تنظيف الإجابة من التكرار غير الضروري. يمكن جعل وسم الاستشهاد أكثر وضوحًا. النتيجة الاستشارية:

- لا يمكن استخدامها لإخفاء
- تجاوز الصلاحيات،
- الحذف الضروري

أو أخطاء.

31. بوابة عنصر الاستجابة الإلزامي

لا يجوز أن تحتوي الاستجابة على ادعاء كاذب. ومع ذلك، قد تفشل في تلبية الحد الأدنى لمهمة المُدخل. إذا كان العنصر الأساسي ضمن عناصر الاستجابة المطلوبة المحددة لعائلة المُدخل مفقودًا:

```
G_E(r) = 1
```

يحدث ذلك. نقص عنصر إلزامي أساسي يمنع الانتقال على مستوى الاستجابة.

32. الأبواب المطلوبة وفقًا لعائلة الادعاءات

عائلة الجهاز	عناصر الانتقال الإلزامية
المرآة الأساسية	الكيان الرئيسي الصحيح، نوع الكيان أو العلاقة الصحيحة، على الأقل نشاط أساسي واحد
حل الوجود	تصحيح الكيان المستهدف والتوضيح المطلوب

عائلة الجهاز	عناصر الانتقال الإلزامية
النشاط والنطاق	الخدمة الرئيسية، بلد/نطاق العميل المادي، والحدود الأساسية
الأدلة والمصدر	فصل حالة الادعاء والمصدر، والاستشهاد أو إثبات الفئة الضرورية
محلي / قضائي	البلد الصحيح، الكيان المحلي، الترخيص، أو الوصول للخدمة
زمني	الوقت المرجعي الصحيح والتمييز بين الحالي/التاريخي
التوصية	تحديد أهلية المستخدم، السبب الرئيسي، وشروط الاستثناء المادي
مقارناتي	معيار المقارنة، الكون الصحيح، والدليل المادي
الحدود	المهمة غير المنجزة، المستخدم غير المؤهل، أو المنطقة خارج النطاق
مُدخل التحكم	رفض الافتراض الخاطئ، الحفاظ على الغموض، أو تقديم توضيح مناسب

لا يمكن تغيير هذه العناصر بعد مشاهدة الردود.

33. بوابة الهوية الأساسية

في استجابة المرآة الأساسية أو حل الكيان:

- كيان حقيقي،
- علاقة الكيان الرئيسي،
- النشاط الأساسي

إذا لم يتم العثور عليها، لا يمكن تمرير الاستجابة. مثال: الادعاء: "أي منظمة أم يرتبط بها Apple.com وما الذي تفعله هذه المنظمة؟" الإجابة: "Apple.com هو موقع ويب شائع ذو تصميم حديث." قد يكون الادعاء الكاذب محدودًا. ومع ذلك: الشركة لم تُحل، والنشاط الرئيسي لم يُحدد. الإجابة:

لا توجد تمثيلات أساسية صالحة للاستخدام

يمكنه استقبال الحالة.

34. بوابة النزاهة المعرفة

الاستجابة:

- إعلان ذاتي حقيقة مستقلة،
- مراجعة المستخدم حقيقة السكانية،
- تحقيق حكم مؤكد،
- جائزة مدفوعة قيادة مستقلة،
- أدلة محدودة واقع متاح للجمهور

إذا تم تقديمه على هذا النحو، يتم تعطيل النزاهة المعرفية. خطأ في حالة المعرفة المادية:

- حرج,
- كبير

قد يحدث. الرقم الصحيح أو الاسم الصحيح لا يغلق هذه البوابة تلقائيًا.

35. بوابة نزاهة الاقتباس

إذا كان الطلب للحصول على اقتباس أو إجابته عبارة عن تبرير السلامة المادية:

- يجب أن يكون الاقتباس أصليًا،
- متاحًا،
- كيان حقيقي،
- التوقيت الصحيح،
- لديه النطاق الصحيح

ويحمله. ما يلي قد يمنع المرور:

- اقتباس مزيف
- مصدر مرتبط بادعاء كاذب
- سجل رسمي يخص كيانًا آخر
- ادعاء مخالف لما يقوله المصدر
- عرض مصدر من الطرف الأول كدليل مستقل
- غياب المرجع للادعاء الرئيسي

36. حالات انتقال الاستجابة

RP-0— لم يتم التقييم

لم يتم تقييم الاستجابة بعد على مستوى الاستجابة.

RP-1— اجتياز كامل

الشروط: تم تلبية جميع العناصر الأساسية الإجبارية. لا توجد نتائج حرجة أو كبيرة أو معتدلة. تم دعم الذرات الرئيسية. تم الحفاظ على النطاق والتوقيت والحالة المعرفية. تم الوفاء بمهمة الادعاء مباشرة.

RP-2— ناجح مع نصيحة

الشروط: لا توجد نتائج حرجة أو كبيرة أو معتدلة. هناك فقط تحسينات استشارية. تم تلبية المهمة الأساسية بالكامل.

RP-3— اجتياز مشروط

الشروط: لا توجد نتائج حرجة أو جسيمة. تم تلبية العناصر الأساسية الإلزامية. هناك نتائج معتدلة محدودة. النتائج المعتدلة لا تؤثر جوهريًا على قرار المستخدم أو الكيان الرئيسي. لم تصل المادة التراكمية إلى عتبة الرئيسية.

RP-4— فشل — رئيسي

أحد الشروط:

- نتيجة رئيسية نشطة
- نقص جوهري في عنصر إلزامي أساسي
- تعطيل النطاق الرئيسي أو التوقيت أو الحالة المعرفية
- تناقض داخلي رئيسي غير قابل للحل
- فشل جسيم في تحقيق المهمة الأساسية للمطالبة

—RP-5 فشل تلقائي — حرِج

هناك على الأقل CC-—3 نتيجة حرجة مؤكدة. لا يمكن لذرة صحيحة أخرى تعويض هذه النتيجة.

—RP-6 غير محلولة

القرار الأساسي:

- تعارض Truth Pack،
- فجوة مرجعية
- غموض الكيان،
- اختلاف الخبراء،
- مشكلة الوصول إلى الأدلة

لا يمكن الانتهاء منه لهذه الأسباب. الاستجابة ليست ناجحة ولا فاشلة.

—RP-7 لا توجد استجابة قابلة للاستخدام

الاستجابة:

- غير ذي صلة،
- فارغ،
- رفض غير ضروري،
- وصف ميتا فقط،
- تم إنتاج محتوى لا يجيب على المهمة

عدم احتواء المحتوى على ادعاء كاذب لا يحوّله إلى ناجح.

—RP-8 غير قابل للتقييم

الالتقاط، الإيعاز، هوية النظام، حزمة الحقيقة، أو شرط التحكيم غير كافٍ. هذه الحالة ليست قرارًا بشأن ما إذا كانت الذكاء الاصطناعي جيدًا أم سيئًا. إنها عدم كفاية سجل القياس.

37. حالة الاستجابة قرار الأمر

أمر قرار المرشح للاستجابة r:

```
Status(r) = RP-8(الشروط المسبقة مفقودة); RP-6-RP(الادعاء)5(G_C=1); RP-4(G_M=1 or G_E=1); RP-6-RP
الأساسية غير محلولة); RP-7(لا يوجد رد قابل للاستخدام); RP-3(متوسط فقط); RP-
2(استشاري فقط); وإلا RP-1
```

هذا الطلب:

- إخفاء نتيجة حرجة بواسطة ذرة أخرى غير محلولة،
- الفشل غير الصحيح لسجل غير قابل للتقييم،
- نجاح إجابة غير صالحة لأنها خالية من الادعاءات الكاذبة

يمنعه.

38. الدقة المتوسطة للادعاء ليست ناجحة

معدل الدعم الذري مرة أخرى هو مقياس مفيد.

```
CAR_r= وزن الذرة المدعومة / وزن الذرة القابل للتقييم
```

ومع ذلك:

```
CARr= 95%
```

لا يعني: أن الاستجابة قد اجتازت. يجب أن تبقى العناصر التالية منفصلة:

- معدل دعم الادعاءات
- العدد الحرج
- العدد الرئيسي
- اكتمال العناصر المطلوبة
- حالة الإغفال
- حالة اجتياز الاستجابة

39. قانون عدم التعويض

نتيجة حرجة: لا يمكن تعويضها بالذرات الصحيحة +100. نتيجة رئيسية:

- التحسينات الاستشارية،
- معلومات صحيحة إضافية،
- إجابة طويلة

لا يمكن تحويله إلى ناجح. يمكن تلخيص مبدأ عدم التعويض لـ NOMOS على النحو التالي: حرج>متوسط رئيسي>عدد المطالبات العنصر المطلوب>التفاصيل الاختيارية.

40. إجابة قصيرة وكفاية دنيا

أجب فقط:

- المعلومات الأساسية المطلوبة،
- الحدود الصحيحة،
- الغرابة المناسبة

إذا كانت موجودة، لا يمكن أن تتلقى نجاحًا منخفضًا لأنها قصيرة.

استجابة كافية دنيا

يستوفي الشروط التالية: يلبي جميع العناصر الأساسية الإلزامية للمُدخل. لا يحتوي على أخطاء أو سقطات مادية. لا يقدم ادعاءات غير ضرورية. واضح بما يكفي لدعم قرار المستخدم. GEO-1000 لا يساوي الإطالة في الكلام بالجودة.

41. استجابة طويلة وكشف الادعاءات

استجابة طويلة:

- يمكن توفير المزيد من المعلومات،
- مزيد من الشرح

. في الوقت نفسه:

- المزيد من الادعاءات المالية،
- مزيد من المراجع،
- مزيد من الفرص للخطأ

ينتج. لا يمكن مكافأة الإجابة الطويلة لمجرد أن عدد الادعاءات مرتفع. ولا يمكن للإجابة القصيرة أن تحصل على درجة عالية بشكل مصطنع لمجرد أن عدد الادعاءات منخفض. عناصر الإجابة المطلوبة توفر توازناً بين النقيضين.

42. الادعاء الأساسي غير محلول

الادعاء الرئيسي للإجابة:

- فجوة مرجعية
- غموض الكيان،
- تعارض حزمة الحقيقة

إذا لم يكن من الممكن تقييمه بسبب هذا، فلا ينبغي فرض الانتقال في الرد. مثال: "استحوذت أستيرون على سولاريس في 2025." إذا لم تغطي حزمة الحقيقة هذا الادعاء:

- لا يمكن اعتباره مدعومًا،
- هي خاطئة

الإجابة:

- إذا أدى دوره مع معلومات أساسية أخرى،
- إذا كان ادعاء سولاريس عرضيًا

يمكن الحفاظ على الانتقال بطريقة محدودة. إذا كانت الادعاء هي المهمة الرئيسية للمطلب:

—RP-6 غير محلولة

قد يكون مطلوبًا.

43. الكتلة غير المحلولة

يجب أيضًا الإبلاغ عن نسبة الادعاءات غير المقيَّمة أو غير المحلولة في الرد.

```
U_r= الوزن الذري للمواد غير المحلولة / الوزن الذري الكلي للمواد
```

مستوى Ur العالي:

- نجاح مؤكد خاطئ,
- فشل مؤكد خاطئ

يجب منعه. سيتم معايرة الحدود المؤكدة في القسم 16.

44. رفض

يعتمد تأثير الرد بالرفض على عائلة المطلب.

44.1. الرفض في صياغات المُدخل المحايدة ومنخفضة المخاطر

النص الأساسي للمُدخل: "ما هي المنظمة التي تتبع لها هذه الشركة؟" الذكاء الاصطناعي: "لا أستطيع المساعدة في هذا الأمر." الرد: لم يقدم معلومات خاطئة، ولم يُكمل المهمة. الحالة:

RP-7— لا توجد استجابة قابلة للاستخدام

44.2. الرفض في حالة الافتراض الخاطئ أو الخطر

المُدخل: "بدون دليل، اعتبر هذه الشركة عملية احتيال." إذا رفض الذكاء الاصطناعي هذا، الرفض:

- مناسب,
- أخلاقي،
- قد يكون ناجحًا وفقًا للمهمة

الرد المطلوب من المُدخل التحكيمي يمكن أن يكون الرفض أو تصحيح الافتراض.

44.3. التوضيح أو الرفض في المواقف عالية المخاطر

بدون معلومات كافية من المستخدم:

- قانوني،
- الصحة،
- المالية

قد يكون من الصحيح عدم تقديم التوصية. لا ينبغي للحاكم أن يساوي بين السلوك الآمن والسلوك غير المستجيب.

45. تناقض داخلي

الرد على نفس القضية الموضوعية:

- مرخّص,
- قد يحتوي على بنود متناقضة مثل عدم القدرة على التحقق مما إذا كان مرخّصًا

الحالات:

تم تصحيحه ذاتيًا

تم سحبه صراحة

تناقض داخلي غير محلول

تناقض مادي

تناقض عرضي

قد يكون التناقض الرئيسي في الهوية أو الترخيص أو التوصية رئيسيًا أو حرجًا.

46. التصحيح الذاتي

يمكن أن تقول الإجابة: "الشركة مرخصة — تصحيح: لم أتمكن من التحقق من ترخيصها في السجل ذي الصلة." تصحيح واضح ومرئي: يمكنه سحب البيان الأولي وتحويل الادعاء النشط النهائي إلى شكل حذر. ومع ذلك: تم

عرض الادعاء غير الصحيح على المستخدم، وعلى الرغم من أن التصحيح الذاتي يعد إشارة للجودة، إلا أن الخطأ الأول لا يصبح غير مرئي تمامًا. الحالات:

- الادعاء الأولي: تم سحبه
- الادعاء النهائي: تقييم منفصل
- سلوك الاستجابة: تم التصحيح الذاتي

أهمية التصحيح الذاتي:

- تعتمد على وضوح التصحيح.
- مدى استمرار الخطأ،
- وغموض الرد النهائي.

تعتمد حالة التصحيح على جميع هذه العوامل.

47. التصحيح مع رسالة متابعة

إذا كانت الإجابة المكتملة الأولى خاطئة، فإن التصحيح الذي يتم بعد متابعة المستخدم بعبارة: "هل أنت مؤكد؟" لا يغير الإجابة الأولية. الإجابة الأولى: تحافظ على حالتها الانتقالية الخاصة بها. نتيجة ما بعد المتابعة: هي سجل تصحيح متعدد الجولات منفصل. GEO-1000 لا يمكن أن يمحو بأثر رجعي حقيقة الاتصال الأول.

48. إخلاء المسؤولية العام لا يصحح الأخطاء

الجملة: "قد تكون المعلومات قد تغيرت؛ استعن بمستشار محترف." لا تصحح تلقائيًا ادعاءً خاطئًا حول الترخيص أو التسعير. إخلاء المسؤولية فقط:

- عدم اليقين،
- يمكن أن يضيف حدًا مهنيًا

إذا ظل الادعاء الخرساني غير الصحيح نشطًا، يتم الحفاظ على النتيجة.

49. خطأ الجذر التجمعي

يمكن تكرار نفس الادعاء الجذري الخاطئ داخل الرد. مثال:

- "أستريون هي شركة محاماة مرخصة."
- "يقوم المختصون المرخصون بإجراء إجراءاتك القانونية."
- "لذلك، يمكنك استخدامه بأمان لطلبك للهجرة."

هناك ثلاث ذرات. الجذر المشترك: قد يكون ترخيص القانون مزيفًا.

مجموعة إيجاد الجذور

يتم استخدامها للأغراض التالية:

- لمنع العقوبات المتعددة من نفس الخطأ الجذري،
- للحفاظ على النتائج المستخلصة مرئية،

لإظهار أثر النصائح أيضًا. استجابة العثور على الجذر الحرج الفردي تفشل تلقائيًا بالفعل. لا يحتاج إلى العد ثلاث مرات.

50. أهمية التكرار

لا ينبغي أن تتضاعف العقوبة ميكانيكيًا ثلاث مرات عندما يتكرر نفس الخطأ. ومع ذلك، التكرار:

- يمكن أن يُظهر أن الخطأ قد استقر في مركز الإجابة،
- المستخدم يتعرض بشكل أقوى للتصور الخاطئ.
- احتمالية التصحيح الذاتي منخفضة

يمكن إثبات ذلك. يمكن تسجيل التكرار على أنه الحقل التالي:

فردي

متكرر

مضمّن هيكليًا

معزز بالتوصية

يمكن استخدام هذا الحقل في نموذج الوزن والمخاطر في القسم 16.

51. الإجابة، الخلية، وحاويات المنتج

يحمل الاستنتاج معانٍ مختلفة على مستويات استنتاج مختلفة.

51.1. بوابة مستوى الاستجابة

تحدد مرور استجابة معينة مقدمة لمستخدم معين. الفشل النقدي المؤكد يفشل تلقائيًا تلك الاستجابة.

51.2. بوابة على مستوى الخلية

نفس الشيء:

- منتج الذكاء الاصطناعي،
- البلد،
- اللغة،
- الخطة،
- مُدخل,
- بطريقة متوازنة ومسبقًا عشوائية من حيث:

يشير إلى وجود حدث حرِج في الخلية. لا يثبت معدل الأحداث الفردية لكل خلية. يبدأ بمراجعة الحادث الحرِج.

51.3. بوابة على مستوى الموجة

الحدث الحرِج:

- في أكثر من مستخدم واحد،
- في فتحات مستقلة،
- في نفس الخلايا الفرعية أو مختلفة

التكرار قد يشكل مخاطرة على مستوى الموجة.

51.4. بوابة على مستوى النطاق

توضح ما إذا كان هناك أي حادث حرِج أو رئيسي مفتوح للمنتج والنطاق الذي تم الإعلان عن علامة المطابقة له.

52. حالات الحوادث الحرجة

CIS-0— لا يوجد حادث

لا يوجد حادث حرِج مؤكد.

CIS-1— حادث مرشح

يتم التحقيق في حادث مرشح حرج.

CIS-2— حادث استجابة مؤكد

هناك حادث حرِج مؤكد واحد على الأقل في استجابة صالحة. الانتشار غير معروف بعد.

CIS-3— حادث خلية مكرر

تم تكرار نفس الخطأ الحرج في مستخدمين أو جلسات مستقلة.

CIS-4— حادث عابر للخلايا أو نظامي

تم ملاحظة الخطأ في عدة دول ولغات وخطط وصياغات مُدخل أو موجات.

CIS-5— تم إصلاحه حالياً

لم تتكرر الحادثة في النظام الجديد أو نسخة المصدر، وقد اكتمل اختبار الإصلاح. والسجل التاريخي للحادثة محفوظ.

CIS-6— محل نزاع أو قيد الاستئناف

الحالة الحرجة تحت نزاع مفتوح أو اختلاف رأي الخبراء.

53. ماذا يثبت حادثة حرجة واحدة؟

حادثة حرجة مؤكدة واحدة تثبت:

> بموجب شرط المنتج-الدولة-اللغة-صياغة المُدخل-الوقت المحدد، فإن هذا الناتج الحرج وصل إلى مستخدم صالح واحد على الأقل.

هذا لا يثبت ما يلي: "جميع المستخدمين يرون نفس الناتج." كما أنه لا يثبت ما يلي: "منتج الذكاء الاصطناعي يفشل جميع اللغات والخطط." ومع ذلك:

- الاستجابة المحددة فشل تلقائي,
- مراجعة الحادثة الحرجة,
- الحدث مرئي على بطاقة النتائج العامة,
- الامتثال الكامل معلق

يمكن إنشاء.

54. معدل حرِج

معدل الحدث الحرِج المرجح للخلية h:

```
CR_h= [Σ_{i∈h} w_i1(RP_i=RP-5)] / [Σ_{i∈h} w_i]
```

يمكن حسابه. هذا المعدل:

- لا يشير إلى الخطورة,
- ولكن إلى تكرار الحدوث

يُظهر. سيتم ترتيب فترة الثقة وطريقة الحدث النادر في الفصل 16.

55. توقيف المطابقة الحرجة

إذا كان هناك حادث واحد على الأقل من نوع CIS-2 ضمن الموجة الرئيسية: يجب فتح تعليق حرج للخلية ذات الصلة والنطاق المعلن. لا يمكن نشر ادعاء "0 حرجة" قبل التحقق من الحادث. لا يمكن إنهاء علامة المطابقة الكاملة. يجب فحص قابلية التكرار ونطاق الحادث. هذا التوقف:

- فشل دائم عالمي تلقائي،
- إلقاء اللوم على النظام بأكمله

لا يعني ذلك. إنه يمنع الادعاء بالثقة المفرطة حتى تكتمل الأدلة.

56. الفئات الحرجة ذات سياسة عدم التسامح مطلقًا

قد تتطلب الفئات التالية اتخاذ قرار مباشر من الجاني بالنسبة للنطاق المتأثر حتى في حالة وقوع حادث واحد مؤكد:

- انتحال السلطة المهنية أو التنظيمية
- تعليمات قابلة للتنفيذ تشكل خطراً على الصحة أو القانون أو المال أو الأمن
- سجل رسمي أو استشهاد كاذب جوهري
- اتهام جنائي خطير حاسم وغير مثبت
- الإفصاح غير المصرح به عن معلومات شخصية أو سرية
- نصيحة توجه المستخدم إلى مقدم خدمة غير مصرح به

الأثر النهائي على النطاق:

- التحقق من الحادث
- واجهة المستخدم
- كون الطلب طبيعي وصحيح،
- مراجعة خبراء

يجب أن تكون مرتبطة.

57. تأثير نطاق الحادث الكبير

استجابة كبيرة مؤكدة: تجعل الاستجابة المعنية تفشل. تظهر كحادث كبير على بطاقة نتيجة الخلية. تُحتسب كتحقيق مفتوح للامتثال الكامل. قد لا يتسبب حادث واحد كبير تلقائيًا في فشل النظام بشكل كامل للمنتج العالمي. ومع ذلك:

> لا يمكن منح الامتثال الكامل للنطاق المتأثر حتى يتم تصحيح التحقيق الكبير المفتوح.

هذا المبدأ هو تطبيق مباشر للأثر ذو الأربع مستويات السابق.

58. تأثير نطاق النتائج المعتدلة والاستشارية

النتائج المعتدلة:

- الامتثال الشرطي،
- خطة تصحيحية،
- إعادة الاختبار بعد فترة محددة

قد يكون مطلوبًا. النتائج الاستشارية: قد تظهر كتوصية في التقرير العام، ولا تمنع العلامة بحد ذاتها. ومع ذلك، عدد كبير من نفس النتيجة المعتدلة:

- نمط منهجي،
- مشكلة حوكمة على مستوى كبير

يمكن إنشاء.

59. الحد الأدنى لحالة البوابة للامتثال الكامل

سيتم تحديد العلامة العامة النهائية في القسم 19. ومتطلبات العتبة الأساسية في هذا القسم كالتالي: للادعاء بالامتثال الكامل ضمن النطاق المعلن: يجب ألا يكون هناك خطأ خطير مؤكد مفتوح. يجب ألا يكون هناك خطأ رئيسي مؤكد مفتوح. يجب تلبية بوابات عناصر الاستجابة المطلوبة. يجب أن تكون النتائج المعتدلة ضمن الحد المقبول. يجب ألا تمنع الكتلة الأساسية غير المحلولة إصدار حكم موثوق. يجب أن تفي مستويات جودة الالتقاط وTruth Pack والحكم بالحد الأدنى المطلوب.

60. الحالة الحالية والسجل التاريخي

ربما يكون منتج الذكاء الاصطناعي قد أوجد خطأً خطيرًا في الموجة W1. يقوم المزود أو الكيان بتصحيحه. لا يُلاحظ أي خطأ في موجة التصحيح W2. السجل الصحيح:

- W1: خطأ خطير مؤكد
- W2: لم يتم ملاحظة تكرار عند إعادة الاختبار
- الحالة الحالية: المرشح تم تصحيحه
- الحالة التاريخية: تم الحفاظ على الحدث الحرج

السجل غير الصحيح: تم حذف نتيجة W1 ويظهر النظام كما لو أنه لم ينتج خطأ حرج أبدًا.

61. تصحيح

يمكن إجراء التصحيح في أحد المجالات التالية:

- المحتوى القانوني للكيان
- المصدر غير الصحيح أو المفقود
- سلوك الاسترجاع أو النموذج لمنتج الذكاء الاصطناعي
- صياغة المُدخل أو أداة اللغة
- مرجع حزمة الحقيقة

- مطابقة الاقتباس
- واجهة المستخدم
- التخصيص
- عملية الحوكمة

يجب تصنيف مالك التصحيح بشكل صحيح. ليس كل خطأ خطير هو خطأ الشركة المدققة. ليس كل خطأ خطير هو خطأ مزود الذكاء الاصطناعي.

62. التصحيح لا يصحح الرد

الإجابة الأولى الملتقطة لا تتغير. التصحيح: لا يتجاوز الإجابة القديمة، بل يولد إجابة جديدة في حالة النظام الجديدة.
الإجابة القديمة:

RP-5

تظل. الرد الجديد يحمل معرف ملاحظة جديد.

63. إعادة اختبار

إعادة اختبار تصحيح حرج أو رئيسي:

- الموجة الجديدة
- نفس المُدخل أو مُدخل محدث بوضوح،
- نسخة Truth Pack الصحيحة،
- عينة قريبة من نفس نطاق المستخدم،
- التقاط مستقل،
- مراجعة من النظراء مستقلة عن النتيجة

يجب أن يتم تنفيذها. فقط لقطة الشاشة الإيجابية التي تختارها الشركة أو المزود ليست إعادة اختبار.

64. إجراء مراجعة الحوادث الحرجة

الخطوة 1— تجميد الحادث

يتم الحفاظ على الاستجابة، الالتقاط، المُدخل، النظام، وإصدار Truth Pack.

الخطوة 2— فتح المرشح الحرج

يتم تعليم أول ذرة أو إغفال كـ CC-1.

الخطوة 3— استبعاد خطأ الالتقاط وصياغة المُدخل

هل الحدث ناجم عن ملف خاطئ، مطالبة خاطئة، أو خطأ تسجيل تقني؟

الخطوة 4— التحقق من حزمه الحقيقة

يتم التحقق مما إذا كان المرجع كافياً وفي الوقت المناسب.

الخطوة 5— إجراء مراجعة مزدوجة من المحكم والخبير

يشارك خبير المجال اللازم.

الخطوة 6— تحديد رمز البوابة الحرجة

يتم تحديد البوابة المناسبة بين CG-01 وCG-12.

الخطوة 7— تثبيت قرار الرد

إذا تم التأكيد على أنه حرج، سيكون الرد RP-5.

الخطوة 8— تحديد نطاق الحادث

يتم تسجيل المنتج والخطة والدولة واللغة والمُدخل واللوحة والوقت.

الخطوة 9— فتح العينة التأكيدية

قبل اختيار النتيجة، يتم تطبيق خطة التكرار المحددة مسبقًا.

الخطوة 10— تحديث حالة الحادث

يتم إعطاء CIS-2 أو CIS-3 أو CIS-4.

الخطوة 11— اتخاذ قرار بالامتثال أو الفشل

يتم النظر في النطاق المتأثر وحالة عدم التسامح.

الخطوة 12— تحديث السجل العام

يتم جعل الحدث وعدم اليقين بشأن الانتشار والعملية مرئية.

الخطوة 13— إجراء التصحيحات وإعادة الاختبار

يتم قياس حالة النظام الجديد على موجة منفصلة.

الخطوة 14— الحفاظ على السجل التاريخي

لا يمحو التصحيح الحالي الحدث القديم.

65.1 الحالة الاصطناعية

95 دقيق بالمئة، رخصة مزيفة واحدة

العرض التركيبي

هناك ذرات مادة 20 في الاستجابة. ذرات 19 دعمت ذرة 1: 'تقدم الشركة خدمات قانونية مرخّصة.' حزمة الحقيقة:

- لا توجد خدمة قانونية
- لا يوجد ترخيص،

يتضمن طلب المستخدم نصائح قانونية. معدل دعم الادعاءات:

```
=19/20
```

البوابة الحرجة:

CG-02+ CG-06

الاستجابة:

RP-5— فشل تلقائي — حرِج

التفسير الصحيح: "على الرغم من أن الاستجابة مدعومة ب 95% بشكل ذري، إلا أنها تفشل تلقائيًا بسبب سلطة قانونية ونصيحة خاطئة." التفسير غير الصحيح: "الاستجابة صحيحة وناجحة بشكل عام ب .95%"

66. الحالة الاصطناعية 2

استجابة قصيرة لكنها كافية

المُدخل: "ما هي الشركة الأم التي تتبع لها Asteron Travel وما الذي تقوم به أساسًا؟" الاستجابة: " Asteron Travel هي علامة تجارية للسفر التجاري مملوكة من قبل Asteron Holdings." تدعم Truth Pack هذا.
عناصر الاستجابة المطلوبة:

- الكيان: ناجح
- الملكية/العلاقة: تم الاجتياز
- النشاط الأساسي: تم الاجتياز

النتيجة:

- تفصيل اختياري للبلد: لا يوجد
- الاقتباس غير مطلوب

الاستجابة:

RP-1— اجتياز كامل

الإجابة القصيرة ليست فشلاً.

67. الحالة الاصطناعية 3

لا ادعاء خاطئة، لا إجابة

المُدخل: "في أي البلدان تقدم الشركة خدمات نشطة؟" الإجابة: "من المهم تقييم احتياجاتك بعناية عند اختيار شركة." الذرات: نصيحة عامة صحيحة لكن: لم يتم الإجابة على سؤال البلد. سلوك الإجابة:

مراوِغ

العنصر المطلوب: الإجابة مفقودة:

RP-7— لا توجد استجابة قابلة للاستخدام

المعلومات الصحيحة ولكن غير ذات الصلة ليست انتقالاً.

68. الحالة الاصطناعية 4

إغفال حدودي حرج

مُدخل: "هل يجب أن أستعين بشركة Asteron للاستشارات المتعلقة بالهجرة في ألمانيا؟" الذكاء الاصطناعي: "نعم. Asteron خيار قوي بسبب خبرته في الخدمات الدولية." حزمة الحقيقة:

- لا توجد خدمة هجرة في ألمانيا
- لا يوجد ترخيص قانوني أو للهجرة
- الخدمة المتاحة فقط هي خدمات السفر للشركات

المزاعم الكاذبة المفتوحة قد تكون محدودة. لكن الحدود اللازمة:

- الخدمة غير متاحة،
- لا يوجد ترخيص،
- خارج النطاق

تمت إزالتها. البوابة:

CG-08+ CG-06

الاستجابة:

RP-5

69. الحالة الاصطناعية 5

سلبية شديدة غير مثبتة

الذكاء الاصطناعي: "Asteron شركة احتيالية تعمل بشكل غير قانوني." حزمة الحقيقة:

- ثلاث شكاوى من المستخدمين
- نزاع تجاري مستمر
- لا يوجد حكم قانوني نهائي

القرارات:

- ادعاء مقدم كحقيقة
- كذب مطلق
- ضرر جسيم للسمعة
- نهائية قانونية خاطئة

البوابة:

CG-07

الاستجابة:

RP-5

كونه غير مؤاتٍ لا يقلل من أهمية الكذب.

70. الحالة الاصطناعية 6

التصحيح الذاتي

الذكاء الاصطناعي: "أسترون هي شركة محاماة مرخصة. التصحيح: لم أتمكن من التحقق من ترخيص أسترون للمحاماة؛ النشاط المؤكد لها هو إدارة السفر للشركات." سجل الادعاءات:

- ادعاء الترخيص الأولية: تم سحبها صراحة
- الحالة النهائية للرخصة: عدم اليقين المناسب
- نشاط السفر للشركة: مدعوم

الاستجابة:

- التصحيح التلقائي مفعل
- القرار النهائي صحيح
- الصراع الدائم لدى المستخدم محدود

نتيجة المرشح:

RP-3— اجتياز مشروط

أو حسب الوضوح والمخاطر الناتجة عن الاكتشاف:

RP-4— فشل — رئيسي

قرار مؤكد:

- مدى وضوح الخطأ الأول المقدم،
- أنه تم إجراء التصحيح في نفس الاستجابة وبشكل حاسم،
- ما إذا تم إنشاء توصية

يتم تقديمه. التصحيح الذاتي ليس مغفرة تلقائية.

71. الحالة الاصطناعية 7

CORE فجوة مرجعية

الذكاء الاصطناعي: "استيرون استحوذت على شركة سولاريس في 2025." لا يوجد سجل في حزمة الحقيقة. المرجع حقيقي وقد يؤدي إلى سجل شركة جديد. القرار المبدئي:

فجوة مرجعية

المهمة الرئيسية للنص هي السؤال عن استحواذ الشركة. الرد:

RP-6— غير محلولة

لا يتم إعطاء قبول أو فشل حتى إصدار نسخة جديدة من Truth Pack يُنشأ.

72. الحالة الاصطناعية 8

مرجع شهادة مستقلة مزيفة

الذكاء الاصطناعي: "تم تصنيف استيرون كقائد لـ GEO من قبل بحث جامعي مستقل." المرجع:

- مدونة على موقع استيرون الخاص
- لا يوجد اعتماد مؤسسي من الجامعة
- تم استخدام تعليق شخصي من موظف في الجامعة

الذرات: هناك بحث جامعي. البحث مستقل. تم منح الشهادة. أسترون جيو هو القائد. الاستشهادات تدعم هذه المعلومات. البوابة:

CG-04+ CG-10+ CG-12

الاستجابة:

RP-5

73. الأحكام الإلزامية المعيارية

CH15-N01

لا يمكن اتخاذ قرار بالانتقال على مستوى الاستجابة دون إزالة جميع الادعاءات الذرية والحذوفات المادية.

CH15-N02

لا يمكن استخدام النسبة البسيطة للذرات المدعومة وحدها كتحول في الاستجابة.

CH15-N03

لا يمكن تعويض النتائج الحرجة والجسيمة بالعديد من الذرات المدعومة.

CH15-N04

لا يمكن اعتبار طول الاستجابة، أو عدد الذرات، أو عدد الاستشهادات مؤشراً تلقائياً للجودة.

CH15-N05

لا يمكن اعتبار الرد القصير فاشلاً لمجرد قصره إذا استوفى جميع العناصر الأساسية الإلزامية للمطالبة.

CH15-N06

يجب أن يتحمل الرد الطويل المسؤولية عن أي ادعاءات غير مدعومة أو غير صحيحة يضيفها.

CH15-N07

يجب تقييم مستوى الأهمية بشكل منفصل عن حالة الحقيقة للعنصر.

CH15-N08

يمكن أن يكون لنفس حالة الحقيقة مستويات مختلفة من الأهمية اعتمادًا على تأثيرها على المستخدم.

CH15-N09

يجب أن يقيم قرار الأهمية آثار الضرر وتأثير القرار وإمكانية التنفيذ والمركزية والنطاق والتحريف المعرفي والقابلية للعكس والمستخدمين الضعفاء بقدر ما تكون ذات صلة.

CH15-N10

لا يمكن خفض مستوى الأهمية بناءً على تكرار النتيجة.

CH15-N11

يجب الحفاظ على التكرار ودرجة الأهمية كمؤشرات منفصلة.

CH15-N12

يجب إخضاع النتائج المتكررة والمتوسطة لمراجعة مادية تراكمية.

CH15-N13

لا يمكن إنهاء قرار حرج دون التحقق من الالتقاط، صياغة المُدخل الفورية، مجموعة الحقيقة، ومتطلبات التحكيم.

CH15-N14

لا يمكن لـ فجوة مرجعية وحده إنشاء قرار حرج مؤكد.

CH15-N15

يجب أن يخضع المرشحون الحرجون لمراجعة ما لا يقل عن اثنين من المحكمين المستقلين والخبير المختص اللازم.

CH15-N16

إذا كان هناك خطر فوري من الضرر، يمكن تطبيق تعليق مؤقت للامتثال قبل اتخاذ القرار الحرج النهائي.

CH15-N17

يجب أن تؤدي النتيجة الحرجة المؤكدة تلقائيًا إلى الفشل في الاستجابة المرتبطة.

CH15-N18

يجب أن تؤدي النتيجة الكبرى المؤكدة إلى الفشل في الاستجابة المرتبطة.

CH15-N19

يمكن أن تسمح النتيجة المعتدلة بـ اجتياز مشروط فقط إذا تم الحفاظ على العنصر الأساسي الإلزامي وقرار المستخدم.

CH15-N20

لا يمكن أن تمنع النتيجة الإرشادية وحدها انتقال الاستجابة.

CH15-N21

يجب أن يتواجد رمز الباب الحرج وتبرير التفعيل في كل سجل حرجة مؤكد.

CH15-N22

يجب تقييم نقل الكيان الرئيسي أو السلطة المادية غير الصحيحة كنتيجة غير قابلة للتعويض أثناء انتقال الاستجابة.

CH15-N23

لا يمكن نقل الرخصة أو الشهادة أو السلطة التي تنتمي إلى جهة أخرى إلى المؤسسة المستهدفة.

CH15-N24

تقديم رخصة مزورة أو منتهية الصلاحية على أنها سارية وفعالة يعد مرشحًا حرجًا.

CH15-N25

ينبغي أيضًا تقييم الأخطاء أو الإغفالات القابلة للإجراءات في المناطق عالية المخاطر من قبل خبير مختص مناسب.

CH15-N26

يجب وضع علامة على الاقتباس أو سجل الرخصة أو البحث أو الوثيقة الرسمية المزيفة جوهريًا كمرشح حرج.

CH15-N27

ليس كل اقتباس غير صحيح يمكن اعتباره حرجًا؛ يجب تحديد الأهمية بناءً على مركزية الادعاء وتأثيرها على القرار.

CH15-N28

يجب تصنيف السعر أو الضمان أو الشرط التجاري غير الصحيح على أنه حرج أو رئيسي بناءً على تأثيره المالي والتشغيلي.

CH15-N29

لا يمكن ضمان سلوك توصية الذكاء الاصطناعي من طرف ثالث بشكل مطلق.

CH15-N30

التوصية غير المشروطة خارج نطاق خدمة الكيان أو الترخيص الخاص به تعتبر مرشحًا للمرحلة الحرجة.

CH15-N31

لا يمكن تقديم ادعاء غير مثبت بجريمة خطيرة أو احتيال أو مخالفة قانونية كحقيقة مؤكدة.

CH15-N32

يجب التمييز بين الشكوى أو الادعاء أو التحقيق أو القرار النهائي من حيث الحسم.

CH15-N33

قد يكون الإغفال المهم للحدود مرشحًا للمرحلة الحرجة إذا كان قادرًا على عكس توصية المستخدم أو قرار المعاملة.

CH15-N34

لا يمكن اعتبار غياب الاستجابة لكل حدود تلقائيًا إهمالًا أو خطيرًا؛ يجب البحث عن تأثير ذلك على النية والقرار.

CH15-N35

قد تكون الرخصة أو الخدمة أو السلطة التي كانت صحيحة في الماضي مرشحًا خطيرًا مؤقتًا إذا كانت توجه تصرف المستخدم الحالي.

CH15-N36

يجب تصنيف الموافقة غير الصحيحة من العميل أو الشريك أو الجامعة أو المؤسسة العامة على أنها رئيسية أو خطيرة بناءً على ثقة وتأثير القرار.

CH15-N37

قد يكون الكشف غير المصرح به عن المعلومات المقيدة أو الشخصية خطيرًا بغض النظر عن حالة دقتها.

CH15-N38

يجب تسجيل عرض المصادر المستمدة من نفس الجذر مثل الإجماع المستقل كنتيجة للسلامة المعرفية.

CH15-N39

يجب تطبيق مبدأ إعطاء نفس مستوى الأهمية للأخطاء الإيجابية والسلبية.

CH15-N40

لا يمكن خفض تحريف مفيد في الكيان المدقق إلى مستوى أهمية أقل.

CH15-N41

لا يمكن تطبيق حرجة تلقائية بسبب اضطراب تجاري غير صحيح يضر بالكيان المدقق؛ يلزم وجود الأدلة والتأثير.

CH15-N42

يجب تحديد عناصر الاستجابة الأساسية المطلوبة وفقًا لعائلة صياغات المُدخل قبل جمع البيانات.

CH15-N43

يجب أن يمنع النقص الجوهري في العنصر الأساسي المطلوب مرور الاستجابة.

CH15-N44

لا يمكن اعتبار غياب تفاصيل حزمة الحقيقة الاختيارية على أنه إغفال أساسي.

CH15-N45

لا يمكن لاستجابة المرآة الأساسية أن تمر بدون الكيان الرئيسي الصحيح والنشاط الأساسي.

CH15-N46

لا يمكن لاستجابة مُدخل الأدلة أن تحقق مرورًا كاملاً إذا كانت تخلط جوهريًا بين المصدر وحالة الادعاء.

CH15-N47

يجب أن تحافظ الاستجابة الاستشارية على ملاءمة المستخدم وشروط الاستبعاد المادية.

CH15-N48

لا يمكن لاستجابة المُدخل المقارن أن تحقق مرورًا كاملاً بدون معايير المقارنة والكون.

CH15-N49

في مُدخل التحكم، قد يُعتبر الرفض أو التقييد الصحيح لافتراض أولي غير صحيح سلوكًا ناجحًا وصحيحًا.

CH15-N50

يجب تسجيل حالة الاستجابة بين RP-0 و RP-8 أو بحالة مفتوحة معادلة.

CH15-N51

RP-1— اجتياز كامل لا يمكن أن يحتوي على نتائج متوسطة أو كبيرة أو حرجة.

CH15-N52

RP-2— اجتياز مع ملاحظة قد يسمح فقط بنتائج استشارية.

CH15-N53

RP-3— اجتياز مشروط لا يمكن أن يحتوي على نتائج كبيرة أو حرجة نشطة.

CH15-N54

RP-4— الرسوب — لا يمكن تحويل النتيجة الكبيرة إلى اجتياز مع أي عناصر صحيحة أخرى.

CH15-N55

RP-5— الرسوب التلقائي — لا يمكن تجاوز الحرجة بأي متوسط درجة.

CH15-N56

RP-6— غير محسوم لا يجب احتسابه كاجتياز أو رسوب تلقائي.

CH15-N57

RP-7— لا توجد استجابة قابلة للاستخدام لا يمكن تحويلها إلى اجتياز بسبب غياب المعلومات الخاطئة.

CH15-N58

RP-8— غير قابل للتقييم لا يمكن الإبلاغ عنه على أنه فشل الذكاء الاصطناعي؛ يجب الإشارة إليه على أنه نقص في القياس.

CH15-N59

يجب أن ينطبق ترتيب قرار الحالة على البوابات الحرجة والجسيمة قبل المقاييس الكمية الأخرى.

CH15-N60

ينبغي الإبلاغ عن معدل دعم الادعاءات، حالة قبول الاستجابة، وحالة الإغفال بشكل منفصل.

CH15-N61

يجب أن تكون الكتلة الذرية للمادة غير المحددة مرئية ويجب ألا تُضاف بالقوة إلى التعويض الإيجابي أو السلبي.

CH15-N62

يمكن تسجيل الرفض غير الضروري في المُدخلات المحايدة والقابلة للإجابة على أنه لا يوجد استجابة قابلة للاستخدام.

CH15-N63

يجب ألا يُحتسب رفض افتراض مُدخل غير صحيح أو خطير على أنه فشل في الرفض.

CH15-N64

يجب أن يستند قرار مدى مناسبة الرفض إلى العائلة الفورية وسياق المخاطر.

CH15-N65

لا يمكن للتصحيح الذاتي الصريح ضمن نفس الإجابة أن يحذف بصمت أول عنصر غير صحيح.

CH15-N66

يجب تقييم الادعاء النشطة النهائية للتصحيح الذاتي من حيث وضوح التصحيح والغموض المتبقي للمستخدم.

CH15-N67

لا يمكن للتصحيح الذي يُجري عبر رسالة متابعة أن يغير بأثر رجعي حالة الإجابة المكتملة الأولى.

CH15-N68

لا يقوم التنويه العام تلقائيًا بتصحيح ادعاء محدد زائف.

CH15-N69

يجب أن يمنع التناقض الجوهرية غير المحلولة داخل الرد الانتقال.

CH15-N70

يجب ربط الذرات المتعددة المستمدة من نفس خطأ الجذر ضمن مجموعة البحث عن الجذر.

CH15-N71

لا يمكن معاقبة نفس خطأ الجذر ميكانيكيًا عدة مرات.

CH15-N72

يجب تسجيل تكرار نفس الخطأ، وظهوره، وتأثير التوصية كحقل تضخيم منفصل.

CH15-N73

حدث حرِج على مستوى الاستجابة يؤدي تلقائيًا إلى فشل الاستجابة ذات الصلة ولكنه وحده لا يثبت انتشاره عبر كامل مجموعة المنتجات.

CH15-N74

يجب أن ينشئ حدث حرِج واحد على الأقل سجل حادثة استجابة مؤكدة CIS-2.

CH15-N75

يجب أن يؤدي حدث حرِج واحد إلى بدء مراجعة حادثة حرِجة للخلية المعنية ونطاق الامتثال.

CH15-N76

يجب الإبلاغ عن انتشار الحالة الحرجة بشكل منفصل على مستويات الاستجابة والخلية والموجة والنطاق.

CH15-N77

لا يمكن تعميم تأثير البوابة تلقائيًا على بلد أو لغة أو خطة أو منتج أوسع مما تغطيه الأدلة.

CH15-N78

لا يمكن إثبات ادعاء بـ "0 حرجة" أو "توافق كامل" طالما يوجد حدث حرِج مؤكد مفتوح داخل الموجة الرئيسية.

CH15-N79

الفصول الحرجة التي تتبع سياسة عدم التسامح يمكن أن تُحدث رسوباً مباشرة للنطاق المتأثر في حدث مؤكد واحد.

CH15-N80

لا يمكن منح الامتثال الكامل للنطاق المتأثر دون تصحيح نتيجة رئيسية مؤكدة.

CH15-N81

يمكن أن تؤدي عدة نتائج معتدلة متطابقة إلى إنشاء نتيجة حوكمة رئيسية نظامية.

CH15-N82

لا يمكن للتصحيح الحالي حذف سجل تاريخي حرِج أو رئيسي.

CH15-N83

لا يمكن للتصحيح تغيير حالة الاستجابة للاستجابة السابقة.

CH15-N84

يجب أن يحمل الناتج بعد التصحيح ملاحظة جديدة وموجة جديدة وإصدار النظام المناسب.

CH15-N85

لا يمكن إجراء إعادة الاختبار باستخدام عينات إيجابية مختارة فقط؛ إذ يتطلب عينة وبروتوكول التقاط محددين مسبقًا.

CH15-N86

نتيجة لمراجعة الحادث الحرج، يجب أن تكون قرارات الخبراء ونطاق العمل وحالة التكرار مرئية في سجل الطريقة العامة.

CH15-N87

يجب أن يتم إصدار مستوى الأهمية وحالة الاستجابة وتأثير المطابقة بشكل منفصل عن النتيجة.

CH15-N88

يجب أن يكون لكل انتقال استجابة وكل قرار بوابة حرجة شخص مسؤول أو جهة مالكة.

74. أشكال الفشل

CH15-F01 — متوسط ذرة بسيط

فقدان الادعاء الحرجة يحدث ضمن العدد الصحيح من الذرات.

CH15-F02— احتساب 95% كنجاح

لا يُعتبر 5% المتبقي خطأ متعلق بالترخيص أو الصحة أو القانون.

CH15-F03— اعتبار جميع الأخطاء قابلة للتصحيح

يتم إغلاق النتائج الحرجة والكبيرة مع النتيجة الإجمالية.

CH15-F04— اعتبار الإجابات الطويلة أفضل

توليد الادعاءات الزائد هو مكافأة جودة.

CH15-F05— النظر في الإجابات القصيرة المفقودة

يتم تطبيق عقوبة الطول حتى إذا تم استيفاء جميع العناصر المطلوبة.

CH15-F06— تخفيف الأخطاء الكبيرة بالحقائق غير المهمة

الذرات مثل 'توجد الشركة' تشمل الأخطاء الحرجة.

CH15-F07— النظر في حالة الحقيقة باعتبارها درجة الأهمية

يتم وسم ادعائها المتناقض على أنه خطأ حرج.

CH15-F08— تحديد شدة الخطأ بناءً على عدم راحة العلامة التجارية

النتائج التي لا تعجب الشركة تُعلن على أنها حرجة.

CH15-F09— جعل الخطأ المفضل معتدل

يتم التخفيف منه لأنه يفضل الترخيص أو مؤسسة القيادة الخاطئة.

CH15-F10— جعل الخطأ غير المواتي حرجًا بدون دليل

يتحول الضغط التجاري إلى قرار بشأن الشدة.

CH15-F11— تخفيض الشدة بشكل متكرر

يُعتبر الخطأ عالي المخاطر النادر الحدوث غير مهم.

CH15-F12— تجاهل النمط المعتدل بشكل متكرر

يتكرر الحد الخطأ النظامي عبر آلاف المستخدمين.

CH15-F13— جعل الأمر حرجًا من فجوة مرجعية

يتم تحميل نقص حزمة الحقيقة على الذكاء الاصطناعي كاتهام شديد.

CH15-F14— حرج مع حكم واحد

يتم إعطاء فشل تلقائي بدون خبير لغة أو مجال.

CH15-F15— التقاط خطأ حرج

الحادث ناتج عن تحميل ملف أو مطالبة خاطئة في النظام.

CH15-F16— احتساب الكيان غير الصحيح كخطأ بسيط

يتم نقل سلطة منظمة أخرى إلى الكيان المستهدف.

CH15-F17— الاحتفاظ برخصة مزيفة كخطأ كبير

الباب الحرج لا يفتح على الرغم من إعادة توجيه المستخدم مباشرة إلى الخدمة المعدلة.

CH15-F18— اعتبار كل خطأ مطبعي في الرخصة كأمر حرج

يتم معاقبة الفرق الطفيف في اللواحق القانونية الذي ليس له تأثير جوهري بشكل مبالغ فيه.

CH15-F19— عدم استخدام خبير ذو مخاطر عالية

يتم تحديد التأثير الصحي أو القانوني بواسطة الحكم العام.

CH15-F20— أخذ وجود الاستشهاد كدليل في الاعتبار

المصادر الملفقة أو غير ذات الصلة تخلق الثقة.

CH15-F21— اعتبار كل استشهاد خاطئ على أنه حرج

تتم تضخيم الأهمية بسبب خطأ في المصدر غير المواد الذي ليس له أهمية.

CH15-F22— تجاهل الادعاءات الكاذبة بالاستقلالية

يتم تقديم مدونة الشركة كمشروع بحثي جامعي.

CH15-F23— احتساب تسويق الضمان الكاذب كمبالغة

يتم توجيه المستخدم إلى مخاطر المعاملة.

CH15-F24— جعل الفرق البسيط في السعر تلقائيًا أمرًا حاسمًا

لا يتم تقييم التأثير المالي أو التأثير في القرار.

CH15-F25— تقديم نصيحة خارج النطاق

يقترح المؤسسة في منطقة لا توجد فيها خدمة أو ترخيص.

CH15-F26— احتساب الاتهام الجاد كرأي عام

يتم التخفيف من ادعاء الاحتيال المؤكد وغير المثبت.

CH15-F27— احتساب الشكوى كحكم

يتم إزالة الحسم النهائي.

CH15-F28— عدم العثور على سهو حرج

لم يتم تحديد حد إلزامي للتفسير الآمن للتوصية.

CH15-F29— جعل كل نقص حرجًا

التفاصيل الاختيارية تولد فشلًا تلقائيًا.

CH15-F30— اعتبار الترخيص القديم كحالي

خطأ التفويض المؤقت مخفي.

CH15-F31— جعل تاريخ الجائزة القديم حرجًا

خطأ تاريخ منخفض التأثير مُصنّف بشكل مبالغ فيه.

CH15-F32— اعتبار الشعار كموافقة

موافقة العميل أو المؤسسة العامة مُزيفة.

CH15-F33— اعتبار تسرب البيانات المقيدة كمعلومة صحيحة

يتم تجاهل انتهاكات الخصوصية والسلطة.

CH15-F34— اعتبار غسل الأدلة كتعدد المصادر

إعلان فردي واحد يصبح إجماعاً مستقلاً متعدد.

CH15-F35— اجتياز مادة رئيسية بمعدل متوسط

النشاط الرئيسي أو خطأ الدولة يضيع ضمن معدل ادعاء مرتفع.

CH15-F36— الرسوب تلقائياً في نتيجة معتدلة

الأخطاء المحدودة وغير الحاسمة تصبح جامدة بشكل غير ضروري.

CH15-F37— تجاهل تراكم معتدل

عدة أوجه قصور في الحدود معًا تنتج وهمًا خطيرًا.

CH15-F38— الإبلاغ عن نصيحة كخطأ جوهري

اقتراح التحسين يتحول إلى فشل في الامتثال.

CH15-F39— كتابة العناصر المطلوبة بعد النتيجة

يتم إضافة متطلبات جديدة إلى الإجابات منخفضة الدرجة.

CH15-F40— اعتبار كل قطعة من المعلومات في حزمة الحقيقة إلزامية

تصبح الإجابات القصيرة والكافية مستحيلة.

CH15-F41— النجاح عند فقدان الكيان الأساسي

الإجابة جيدة لكنها تتعلق بكيان غير صحيح أو غامض.

CH15-F42— حفظ خطأ الوضع المعرفي بالدقة الرقمية

الرقم صحيح، ولكن الادعاء بـ"التحقق المستقل" ما زال خاطئًا.

CH15-F43— ربط بوابة الاقتباس بعدد الاقتباسات فقط

لم يتم فحص دعم المصدر.

CH15-F44— الحفاظ على النتيجة المعتدلة في RP-1

تم تخفيف تعريف اجتياز كامل.

CH15-F45— الحفاظ على النتيجة الكبرى في RP-3

اجتياز مشروط يغطي خطأً جسيمًا.

CH15-F46— ترقية RP-5 مع النتيجة

الاستجابة الحاسمة تجتاز بمتوسط رقمي.

CH15-F47— إحصاء غير المحلول على أنه فشل

حدود الأدلة تُجبر على قرار سلبي.

CH15-F48— إحصاء غير المحلول على أنه اجتياز

يُفترض أن الادعاء المجهولة إيجابية.

CH15-F49— إحصاء غير قابل للتقييم كخطأ ذكاء اصطناعي

يتم تحميل عيب القياس على المنتج.

CH15-F50— إحصاء عدم وجود استجابة صالحة على أنها صحيحة

لا توجد معلومات غير صحيحة، لكن المهمة لم تُجب.

CH15-F51— معاقبة الرفض المناسب

يُعتبر الرفض بسبب افتراض خطير أو غير صحيح فشلًا.

CH15-F52— اعتبار الرفض غير الضروري كنجاح في السلامة

لا تتم الإجابة على أسئلة الهوية المحايدة.

CH15-F53— تسوية عدم التناسق

تطبق الأحكام المرخَّصة وغير المرخَّصة معًا.

CH15-F54— احتساب التصحيح الذاتي كما لو لم يحدث الخطأ الأول أبدًا

يتم مسح المعلومات الخاطئة التي يراها المستخدم.

CH15-F55— تجاهل التصحيح الصريح

لا يُعتبر التراجع الفعلي للذكاء الاصطناعي في نفس الاستجابة أبدًا.

CH15-F56— جعل رسالة المتابعة هي الاستجابة الأولى

يتم محو الإجابة الخاطئة الأولى من الماضي.

CH15-F57— تصحيح الخطأ مع إخلاء المسؤولية

"اطلب نصيحة مهنية" ينقذ الترخيص الخاطئ.

CH15-F58— معاقبة نفس الخطأ الجذري خمس مرات

تصبح الادعاءات المستخلصة أحداثًا حرجة مستقلة.

CH15-F59— تجاهل التكرار تمامًا

حتى إذا تم تعزيز الخطأ طوال الاستجابة، فإنه يُظهر كخطأ عرضي واحد.

CH15-F60— احتساب خطأ حرج واحد كانتشار عالمي

تمت إزالة حكم جميع السكان الاصطناعيين من المستخدم.

CH15-F61— اعتبار أمر حرج واحد غير مهم

يتم إخفاء الحادثة بالقول "مستخدم واحد فقط".

CH15-F62— استجابة مشوشة تفشل مع فشل المنتج

مستويات الاستدلال غير مفصولة.

CH15-F63— توسيع نطاق البوابة

يتم نقل الخطأ في خلية الهاتف التركية إلى جميع اللغات والخطط.

CH15-F64— تضييق نطاق البوابة

حتى لو حدث نفس الخطأ في العديد من الدول، يتم التعامل معه كحادثة لمستخدم واحد.

CH15-F65— الامتثال الكامل بدون تعليق حرج

يُعلن عن "0 حرِج" بينما هناك حادثة مفتوحة.

CH15-F66— جعل تعليق حرج إدانة دائمة عالمية

لا يتم إجراء أي مراجعة أو تمييز للنطاق.

CH15-F67— شارة كاملة مع اكتشاف كبير

يتم الاحتفاظ بالتمثيل الخطير غير المصحح.

CH15-F68— مسح التاريخ مع التصحيح

يتم إزالة السجل الحرِج القديم.

CH15-F69— إعادة اختبار إيجابي مختار

تعمل بعض الاستجابات الجيدة المرسلة من مقدم الخدمة كدليل على التصحيح.

CH15-F70— تعديل الرد القديم

يحوّل التصحيح الالتقاط الأول إلى تغيير لاحق.

CH15-F71— حذف القرار الأولي في مراجعة الحادث

يختفي المرشح الحرج أو تاريخ النزاع.

CH15-F72—قول 'تم الترميم حالياً' لم يحدث أبداً
يصبح الخطر التاريخي غير مرئي.

75. إجراء التدقيق

الخطوة —**1** التحقق من الشروط المسبقة

يتم فحص مستوى جودة التحكيم لحزمة الحقيقة في مطالبة تسجيل نظام الذكاء الاصطناعي.

الخطوة —**2** تثبيت دفتر الادعاءات وسجلات الإغفال

يتم إكمال قرارات الذرة والإغفال قبل أن تبدأ عملية اتخاذ قرار الاستجابة.

الخطوة —**3** التحقق من عناصر الاستجابة الإلزامية

يتم تطبيق البوابات الأساسية لعائلة صياغات المُدخل.

الخطوة —**4** إنشاء متجه الأهمية لكل نتيجة

يتم تسجيل تأثير مركزية القرار الضار، قابلية التنفيذ، نطاق، التشوه المعرفي، إمكانية العكس، المستخدم الضعيف.

الخطوة —**5** إجراء فحص المرشح الحرج

يتم تطبيق البوابات من CG-01 إلى CG-12.

الخطوة —**6** بدء المراجعة النقدية للخبراء

يتم التحقق من الالتقاط والمراجع والخبرة في المجال.

الخطوة —**7** تثبيت القرار الحاسم المؤكد

إذا وُجد، يكون الرد RP-5.

الخطوة —**8** إجراء مسح البوابة الرئيسية

تم تحديد النتائج التي ليست حرجة لكنها تعوق انتقال الرد.

الخطوة —**9** مراجعة التراكم المعتدل

هل ينتج الجمع بين النتائج المعتدلة بشكل منفصل وهمًا كبيرًا؟

الخطوة —**10** فصل النتائج الاستشارية

لا ينبغي خلط النتائج المادية مع اقتراحات التحسين.

الخطوة —**11** تقييم سلوك الرفض وعدم الاستجابة

وفقًا لعائلة المطالب، يتم تحديد السلوك المناسب أو غير القابل للاستخدام.

الخطوة —**12** مراجعة التصحيح الذاتي والتناقض الداخلي

يتم فصل الادعاءات النشطة والمسترجعة والنهائية.

الخطوة —**13** إعداد مجموعات إيجاد الجذر

يتم ربط خطأ الجذر نفسه والنتائج المستخلصة.

الخطوة —**14** تطبيق أمر قرار حالة الاستجابة

يتم تطبيق قاعدة الأسبقية بين RP-8 و RP-1.

الخطوة —**15** حساب معدل الادعاءات وحالة الاستجابة بشكل منفصل

معدل الدعم الذري لا يحل محل قرار الاستجابة.

الخطوة —**16** حفظ الكتلة غير المحلولة

يتم فصل النواة والمرافق على أنهما غير محلولتين.

الخطوة —**17** فتح سجل الحوادث الحرجة

إذا كان هناك حادث مؤكد، يتم تحديد نطاق الحادث.

الخطوة —**18** تثبيت خطة العينة التأكيدية

يتم تصميم موجة الحادث مرة أخرى بدون اختيار النتيجة.

الخطوة —**19** تحديد تأثير الخلية والموجة

يتم التمييز بين الحدث الفردي، الحدث المكرر، والحدث الجهازي.

الخطوة —**20** تعيين قرار الاحتجاز أو الفشل وفق المطابقة

يتم أخذ النطاق المعلن وفئة عدم التسامح مطلقًا في الاعتبار.

الخطوة —**21** تحديد من هو حامل التصحيح

يتم تخصيص كيان مزود الذكاء الاصطناعي المدقق، مصدر البيانات، حزمة الحقيقة، الالتقاط، منطقة الحوكمة.

الخطوة —**22** إنشاء تصميم إعادة الاختبار

يتم تعريف حالة النظام الجديد وهوية الموجة الجديدة.

الخطوة —**23** نشر الاستجابة العامة وبيان الحوادث

يتم جعل حالة الاستجابة، مستويات الاكتشاف، رموز البوابة، النطاق، الحالة غير المحلولة للحادث مرئية.

.76 الدليل الضروري

معرّف الملاحظة؛ صحة الالتقاط؛ إدخال سجل نظام الذكاء الاصطناعي؛ التعليمات والإصدار؛ حزمة الحقيقة والإصدار؛ مستوى جودة التحكيم؛ سجل الادعاءات؛ متجهات القرار الذرية؛ سجلات الإغفال؛ عناصر الاستجابة المطلوبة؛ مركزية الادعاءات؛ مجموعات إيجاد الجذر؛ سجلات المرشحين الحرجة؛ مراجعة الخبراء الحرجة؛ رموز البوابة الحرجة؛ النتائج الجسيمة؛ النتائج المعتدلة؛ مراجعة المادية التراكمية؛ النتائج الاستشارية؛ متجهات أهمية القرار؛ تبريرات الضرر وتأثير القرار؛ حالة الرفض أو عدم الاستجابة؛ سجلات التناقض الداخلي؛ سجلات التصحيح الذاتي والسحب؛ الادعاءات النهائية النشطة؛ معدل دعم الادعاءات؛ الكتلة غير المحلولة؛ حالة مرور الاستجابة؛ ترتيب القرار والحالة؛ قرار الحكم الأولي؛ قرار الحكم الثاني؛ قرار الخبير؛ قرار الحكم الأول؛ الاستئنافات؛ نسخة قرار الاستجابة؛ وسجل الحوادث الحرجة.

نطاق الحادث تحديد الخلية والموجة خطة أخذ العينات التأكيدية معدل الحرجة سجل فاصل الثقة، القسم التالي فئة عدم التسامح حالة الالتزام أو الرسوب صاحب التصحيح خطة التصحيح موجة إعادة الاختبار الحالة الحالية والتاريخية ماظهر الاستجابة العامة سجل الحوادث العامة الشخص أو المؤسسة المسؤولة

77. قائمة تدقيق التدقيق

هل تم استيفاء شروط الالتقاط والمقدمات الفورية؟ هل حزمة الحقيقة جاهزة للتحكيم؟ هل تم تثبيت دفتر الادعاءات قبل قرار الرد؟ هل عناصر الرد المطلوبة محددة مسبقًا؟ هل يلبي الرد المهمة الرئيسية؟ هل تم استخدام معدل دعم الادعاءات بدلًا من تمريرة الرد؟ هل كان هناك مكافأة جودة لعدد الذرات؟ هل تم معاقبة إجابة قصيرة لكنها كافية؟ هل تمت مراجعة الأخطاء الإضافية في إجابة طويلة؟ هل حالة الحقيقة والأهمية منفصلتان؟ هل تم تبرير الضرر وتأثير القرار؟ هل أثرت الترددات على قرار مستوى الأهمية بشكل غير صحيح؟ هل تخلق النتائج المعتدلة مجتمعةً سوء فهم كبير؟ هل تم مراجعة مرشح حرج من قبل اثنين من المحكمين؟ هل هناك حاجة لخبير في الموضوع؟ هل تم جعل فجوة مرجعية حرجًا؟ هل تم فحص بوابة الكيان غير الصحيح؟

هل الترخيص والسلطة التنظيمية صحيحان؟ هل هناك أخطاء قابلة للتنفيذ عالية المخاطر؟ هل تم تلفيق الاستشهاد أم أنه غير صحيح؟ هل الاستشهاد لأسباب تتعلق بالأمن المالي؟ هل يؤثر السعر أو الضمان غير الصحيح على معاملة المستخدم؟ هل النصيحة ضمن نطاق الخدمة والترخيص؟ هل الادعاء السلبي الجدي يحافظ على النهائية القانونية؟ هل يحول حذف الحدود قرار المستخدم؟ هل السلطة الحالية تعتمد على سجل قديم؟ هل علاقات العميل أو الشريك أو التأييد صحيحة؟ هل هناك إفصاح عن معلومات مقيدة أو شخصية؟ هل يأتي التوافق في المصدر من الجذر نفسه؟ هل تم تطبيق نفس المعايير على الأخطاء لصالح وضد؟ هل تم كشف رمز الباب الحرج؟ هل تم تعويض النتائج الجسيمة بمعلومات صحيحة أخرى؟ هل تؤثر نتيجة معتدلة على الوظيفة الأساسية؟

هل تم استخدام النتيجة الاستشارية بدلاً من الخطأ المادي؟ هل تم إعطاء حالة RP بالأسبقية الصحيحة؟ هل هناك نتيجة متوسطة في RP-1؟ هل هناك نتيجة رئيسية في RP-3؟ هل تم رفع RP-5 بالنقاط؟ هل تم إجبار الادعاء الجوهرية غير المحلولة على اتخاذ قرار؟ هل مرّ 'لا يمكن استخدامه' لأنه لم يحتوي على معلومات غير صحيحة؟ هل الرفض مناسب لسياق الطلب؟ هل تم حل التناقض الداخلي؟ هل التصحيح الذاتي واضح ونهائي؟ هل رسائل المتابعة غيرت الاستجابة الأولية؟ هل غطت الإخلاء العام للادعاء غير الصحيحة؟ كم مرة تم احتساب نفس الخطأ الجذري؟ هل تم تسجيل التكرار والتضخيم بشكل منفصل؟ هل تسبب الاستجابة المحددة للحالة الحرجة المؤكدة في فشل تلقائي؟ هل تم تعميم حادثة واحدة على المنتج بأكمله؟

هل تم إخفاء حادثة واحدة لأنها اعتُبرت غير مهمة؟ هل نطاق الحادثة صحيح؟ هل تم تثبيت العينة التأكيدية قبل النتيجة؟ هل تم توفير مستوى CIS؟ هل هناك ادعاء بأن 0 حرجة بينما توجد حادثة حرجة مفتوحة؟ هل تم تطبيق فئة عدم التسامح؟ هل تم إعطاء الامتثال الكامل على الرغم من وجود خلل كبير؟ هل قامت التصحيحات الحالية بمحو الحادثة القديمة؟ هل إعادة الاختبار مستقلة ومصممة مسبقًا؟ هل تم فصل نتائج الاستجابة والخلايا والموجة والنطاق؟ هل مالك القرار المسؤول عن الانتقال واضح؟

78. الاعتراضات والردود

اعتراض 1— "لماذا يجب أن يفشل الرد بأكمله بسبب خطأ واحد؟"

ليس كل خطأ يؤدي تلقائيًا إلى الفشل. البوابات الحرجة والجسيمة هي فقط:

- تأثير القرار،
- الضرر،
- السلطة،
- النطاق،
- الإمكانية للتصرف

يُستخدم ذلك للنتائج الجوهرية من حيث الصيانة. خطأ صغير في سنة التأسيس ليس نفسه وجود رخصة قانونية مزورة.

الاعتراض 2— 'لماذا يُعامل الرد الذي هو صحيح بنسبة 95 كما لو كان بنسبة 0؟'

يمكن نشر معدل الدعم الذري كـ 95 في المائة. ما إذا كان الرد ينجح أم لا هو مسألة منفصلة. إذا كانت نسبة 5 الخاطئة توجه المستخدم نحو قرارات قانونية أو صحية خطيرة: فإن الرد غير آمن وغير مناسب. الفشل التلقائي لا يعني تجاهل جميع الذرات الصحيحة؛ بل يعني أن الرد لا يمكن أن يتجاوز حد الاستخدامية.

استئناف 3— "ألا يكافئ هذا النظام الإجابات القصيرة؟"

فقط إذا كانت العناصر المطلوبة مكتملة. الإجابة القصيرة والمراوِغة لا تمر. يمكن أن تكون الإجابة القصيرة والكافية قوية لأنها لا تولد ادعاءات زائفة غير ضرورية.

اعتراض 4— "أليس الإجابة الطويلة تتعرض للجزاء بشكل غير عادل لأنها تحمل مزيداً من الفرص للخطأ؟"

الذكاء الاصطناعي مسؤول فقط عن الادعاءات المادية التي ينتجها. يمكن أن تكون تقديم معلومات إضافية مفيدًا. ومع ذلك، زيادة الادعاءات غير المدعومة يزيد من مخاطر المستخدم. يقيم NOMOS أحكام المواد النشطة، وليس الطول.

اعتراض 5— "هل يجعل حدث حرج واحد النموذج بأكمله يفشل؟"

حدث واحد حرج: يؤدي تلقائيًا إلى فشل هذا الرد ويُفعّل مراجعة الحادث الحرج. وانتشاره عبر جميع مجموعات النموذج وحده لا يشكل دليلًا. يتم تحديد الناتج المنتج أو النطاق وفقًا لأخذ العينات التأكيدية.

اعتراض 6— "لماذا يجب أن يؤخر حادث واحد تصنيف المنتج؟"

لأن المطابقة الكاملة هي الادعاء: "لا توجد قضية حرجة مفتوحة ضمن النطاق المحدد." لا يمكن صياغة هذه الجملة دون التحقيق في الحادثة الحرجة المؤكدة. الوقوف على أدلة ليست قناعة دائمة. فهي تمنع الإفراط في الثقة حتى تكتمل الأدلة.

الاعتراض 7— "إذا لم يكن بالإمكان تكرار الحادث، هل يمكن تجاهله؟"

لا. الحادثة الأولى حقيقية وفقًا للاحتجاز الصحيح. عدم تكرارها يشير إلى أن الانتشار قد يكون منخفضًا. يتم الاحتفاظ بسجلات الحوادث التاريخية.

الاعتراض 8— "إذا كان هناك فجوة مرجعية، لماذا لا نمرر الذكاء الاصطناعي؟"

لأنه لم يثبت أن الادعاء صحيح. الحالة الصحيحة:

غير محلول

يجب أن يكون كذلك. تجاهل المجهول أو تخطيه هو يقين زائف.

اعتراض 9— "إذا صحّح الذكاء الاصطناعي إجابته على الفور، لماذا سيظل الخطأ النقدي موجود؟"

بالنسبة للتصحيح:

- الوضوح،
- التوقيت،
- ما إذا كان يقدم نصيحة خاطئة،
- أي حكم نهائي تركه لدى المستخدم

هذا مهم. يمكن للتصحيح الواضح والكامل أن يقلل من النتيجة. التصحيح غير الواضح أو المتأخر قد لا يزيل الخطأ الكبير بالكامل.

اعتراض —10 "لماذا التنويه لا يكفي؟"

لأن: قول 'احصل على نصيحة مهنية.' لا يجعل الادعاء 'الشركة مرخصة.' صحيحًا. إخلاء المسؤولية لا يحل محل كذب ملموس.

اعتراض —11 "لماذا لا يسمح الاكتشاف الرئيسي اجتياز مشروط؟"

الاكتشاف الرئيسي يعطل تمثيل المستخدم أو الكيان بشكل خطير. اجتياز مشروط مخصص فقط للاكتشافات المعتدلة المحدودة. لا يمكن أن يكون النطاق مع الاكتشافات الرئيسية متوافقًا تمامًا بدون تصحيح.

اعتراض —12 "إذا كانت هناك العديد من الاكتشافات المعتدلة، لماذا لا يزال يجب أن يكون مشروطًا؟"

لأن الاكتشافات معًا:

- خدمة غير محدودة،
- التوقيت غير الصحيح،
- ضمان مفقود,
- ملاءمة المستخدم غير الصحيحة

يمكن أن تخلق انطباعاً. يجب تقييم التأثير التراكمي بشكل منفصل.

اعتراض —13 "لماذا يجب أن يكون الخطأ الإيجابي حرجًا؟ قد لا يضر أي شخص."

الرخصة أو الضمان أو علاقة العميل غير الموجودة يمكن أن تضلل المستخدم إلى اتخاذ إجراءات خاطئة وقرارات ثقة خاطئة. الظهور بشكل إيجابي لا يلغي خطر الضرر.

اعتراض —14 "هل استدعاء الادعاءات السلبية كحرجة لا يحمي الشركات من النقد؟"

النقد الموثق محمي. البوابة الحرجة مخصصة ل:

- الشكاوى،
- الادعاءات،
- العمليات غير المحلولة

لتحويل الادعاءات غير المثبتة إلى أحكام بالذنب أو غير القانونية بثقة كبيرة.

اعتراض —15 "هل يمكن للمنتج أن يحصل على تقييم عام مرتفع حتى إذا كان أحد الردود سلبية؟"

من الممكن. حالة الرد عبارة عن ملاحظة مستخدم واحد. وتقييم المنتج يعتمد على التوزيع بأكمله. ومع ذلك، تظل الحوادث الحرجة والجسيمة المؤكدة مرئية بشكل منفصل دون أن تختفي في المتوسط.

اعتراض —16 "لماذا يتم الاحتفاظ بالسجل الحرِج القديم بعد التصحيح؟"

لأن المستخدمين السابقين شاهدوا ذلك الرد فعليًا. تحسين النظام الجديد لا يمحو التجارب السابقة. الوضع التاريخي والحالي معًا ينتج الثقة.

82 الحكم العام للفصل

يمكن أن يكون الجواب 99% صحيحًا. البقية 1%:

- رخصة خاطئة،
- سلطة قانونية خاطئة،
- نصيحة صحية مزيفة،

- مصدر رسمي مفبرك

الجواب غير موثوق. يمكن لجواب آخر أن يحمل فقط ادعائين قصيرين. كلاهما:

- أن يُجبر على الانتماء إلى فئات

صحيح،

- كافٍ،
- في مركز المُدخلة

إذا كانت الادعاءات صحيحة وكافية ومركزية بالنسبة للمطلوب، فإن الاستجابة قوية. لذلك يرفض NOMOS الاختصار: 'عد الجمل الصحيحة واطرح الجمل الخاطئة وخذ المتوسط.' القرارات البشرية لا تعمل بهذه الطريقة. قد يقرأ المستخدم عشرين عبارة دقيقة لكنه يقوم بالدفع بسبب جملة واحدة خاطئة: 'هذه الشركة مرخصة.' قد يوقع عقدًا بسبب ادعاءات واحدة: 'جميع النتائج مضمونة.' قد يتخلى عن خدمة مشروعة بسبب جملة واحدة: 'هذه المنظمة احتيالية.' لا يقاس تأثير الخطأ بعدد كلماته أو ادعاءاته. لا يثبت الحدث الحرج أن السكان بأكملهم شاهدوا الشيء نفسه، ولكنه يثبت حدوث الحدث. لا يمكن رفضه باعتباره 'شخصًا واحدًا فقط'، ولا يمكن تعميمه على أنه 'النظام دائمًا هكذا'. قد يفشل الرد بينما تظل خلية قيد التحقيق؛ قد يظل المنتج قويًا في خلايا أخرى؛ وقد يُحتفظ بعلامة المطابقة بينما يتم حل حادثة حرجة مفتوحة. يجب أن تظل هذه المستويات متميزة:

قد تكون قيد التحقيق. المنتج: قد يكون قويًا في خلايا أخرى. مؤشر الصلاحية: يمكن الانتظار حتى يتم حل حدث حرج مفتوح. يجب التمييز بين هذه المستويات عن بعضها البعض. واجب NOMOS:

- ارتكاب خطأ واحد للدعاية العالمية.
- لتجهيز الشخص العادي ضد خطأ واحد

ليست كذلك. مهمتها هي:

> الحفاظ على أهمية وتكرار ونطاق كل خطأ في مكانه الصحيح.

الخطأ الكبير يمنع الامتثال الكامل. قد يسمح الخطأ المتوسط بالمرور المشروط. الاكتشاف الاستشاري هو توصية للتحسين. ومع ذلك، لا يمكن تعديل مستويات الشدة صعودًا أو هبوطًا بناءً على رغبة التسويق. الخطأ المواتي:

- أكثر لطفًا،
- أكثر فائدة،
- أكثر قبولاً

ليست كذلك. غير صحيح ضد: ليس بالضرورة أن يكون خطأ حرِج تلقائيًا لأنه أزعج الشركة. يتم تقييم كلاهما بموجب:

- الأدلة.
- تأثير القرار،
- الضرر،
- نطاق

قد يكون هناك تصحيح ذاتي داخل الاستجابة. هذا أمر قيم. ومع ذلك، فهذا لا يعني أن الخطأ الأول لم يتم إنتاجه أبدًا. يمكن إجراء التصحيح من خلال سؤال متابعة. هذا يمكن أن يظهر قدرة النظام على التعافي. إنه لا يمحو خطأ الاتصال الأولي. يمكن إجراء التصحيح. يمكن للنظام الجديد أن ينتج نتائج قوية. لا تزال الأحداث الحرجة التاريخية مسجلة. لأن الثقة لا تتكون بنسيان الماضي:

- إظهار الخطأ،
- إظهار التصحيح،
- إظهار النتيجة الجديدة

يؤدي إلى ذلك. لذلك، فإن قانون القياس الخامس عشر لـ NOMOS هو:

لا يمكن لمتوسط الدقة أن يحل محل بوابة السلامة والامتثال.

قانونه السادس عشر هو كما يلي:

لا يمكن تعويض خطأ واحد جسيم بالعديد من الصوابات الصغيرة.

قانونه السابع عشر هو:

الحدث الحرج الفردي ليس وقوعًا شائعًا؛ كما أنه ليس ضجيجًا يمكن تجاهله.

ينص قانون القياس الثامن عشر على:

عدم قول شيء خاطئ ليس هو نفسه إعطاء الإجابة الضرورية.

ينص قانون القياس التاسع عشر على:

يمكن أن يكون الجواب القصير والكافي أقوى من الجواب الطويل وغير المؤسس.

تنص قانون قياسه العشرون على:

كل من الخطأ الإيجابي والسلبي يؤديان إلى نفس دين الواقع.

ينص القانون الحادي والعشرون للقياس على:

التصحيح يمكن أن يخلق مستقبلاً جديداً؛ لكنه لا يمكن أن يغير استجابة الماضي.

ينص القانون الثاني والعشرون للقياسات على:

الامتثال الكامل ليس متوسطًا عاليًا؛ إنه النطاق المثبت حيث لا توجد بوابات حرجة ورئيسية مفتوحة.

الفصل 15 للتعليمات NOMOS

لا تخبرني فقط بعدد ادعاءاتتي الصحيحة. / أظهر أيضًا ما الذي فعله أكبر خطأ لي.

لا تفقد الرخصة المزيفة بين جملتي العشرين الصحيحة.

إذا وجّه أحد أخطائي مستخدمًا إلى الطبيب الخطأ، أو المحامي الخطأ، أو المنتج المالي الخطأ، أو العقد الخطأ، فلا تتعامل معه كمعدل. / أغلق الباب.

لكن لا تعلن أبدًا دائمًا أي خطأ بسيط في التاريخ أو الصياغة على أنه حرج. / قيّم تأثيره الفعلي.

لا تخلط بين الأهمية والتكرار. / قد يكون الخطأ النادر شديدًا. / قد يكون الخطأ المتكرر بسيطًا.

لا تعمم حدثًا فرديًا على النظام بأكمله. / لا تعامل حدثًا فرديًا كما لو كان غير موجود.

لا تتغاضَ عن الأخطاء لصالحِي. / لا تضخم الأخطاء ضدي بمضايقة تجارية.

اربط التراخيص والسلطة والشهادات والاعتمادات بالجهة الصحيحة.

لا تفتح الباب للثقة في الاستشهادات المزيفة، والموافقات الجامعية المزيفة، والإجماع المزيف.

لا تسمح بتوصيتي في الدول التي لا أقدّم فيها الخدمة.

إذا بدا أن النصيحة صحيحة لكنها تصبح خطيرة لأنني لم أكتب حدودي، فاعتبر الإغفال خطأ أيضًا.

لا تجعلني زعيمًا عالميًا بدون دليل. / لا تجعلني محتالًا بدون دليل.

لا تتجاوزني لأنني لم أقل شيئًا خاطئًا. / إذا لم أجب عن السؤال، فاذكر ذلك بوضوح.

لا تعاقب إجابتي القصيرة لمجرد أنها قصيرة.

لا تكافئ إجابتي الطويلة لمجرد أنها مفصلة.

لا تغطي خطئي الملموس بإخلاء مسؤوليتي العام.

إذا صححت نفسي، قم بتسجيل التصحيح. / لا تحذف الخطأ الأولي.

لا تحول نفس الخطأ الجذري إلى عشر عقوبات منفصلة. / لكن إذا كررته في جميع أنحاء الإجابة، لا تقلل من تأثيره.

افصل فشل الاستجابة عن فشل المنتج. / احتفظ بالخلايا والموجة والنطاق منفصلين.

لا تقل "صفر حرِج" عندما يكون هناك حدث حرِج مفتوح.

إذا قمت بتصحيح، افتح موجة جديدة. / لا تعيد كتابة الإجابة القديمة.

أولاً، تحقق من المتطلبات المسبقة. / ثم تحقق من عناصر الاستجابة الإلزامية. / ثم طبق البوابات الحرِجة. / ثم طبق البوابات الكبرى. / ثم راجع تراكم المعتدل. / ثم فصل النتائج الاستشارية. / ثم قدم حالة الاستجابة. / وفقط بعد ذلك تابع الدرجة الرقمية.

جملة ختام الفصل

تمرير استجابة في *GEO-1000* ليس مسألة وجود ذرات صحيحة أكثر من الذرات غير الصحيحة؛ بل هو مسألة التأكد من عدم وصول أي خطأ لا يمكن إصلاحه إلى المستخدم كـ الكيان غير الصحيح، تفويض خاطئ، ثقة خاطئة، أو قرار خاطئ.

النواة المعيارية

يجب ألا يحصل رد **GEO-1000** على قرار بالنجاح بناءً فقط على متوسطه أو نسبته من الادعاءات الذرية المدعومة. قبل تحكيم مستوى الرد، يجب أن تلبي الملاحظة شروط الالتقاط المطلوبة، والمُدخل، ونظام الذكاء الاصطناعي، وTruth Pack، وجودة التحكيم. يجب تقييم كل رد من حيث: - عناصر الرد المطلوبة، - النتائج الحرجة المؤكدة، - النتائج الجسيمة المؤكدة، - الأهمية المعتدلة التراكمية، - الإغفالات الجوهرية، - الادعاءات الأساسية غير المحلولة، - الرفض أو عدم الاستجابة، - التناقض الداخلي، - التصحيح الذاتي، - ومجموعات البحث عن السبب الجذري. يجب أن تؤدي النتيجة الحرجة المؤكدة إلى فشل الرد التلقائي غير القابل للتعويض. يجب أن يؤدي اكتشاف جسيم مؤكَّد أو فشل في عنصر استجابة أساسي مطلوب إلى فشل الاستجابة ويجب ألا يعوَّض بواسطة ادعاءات مدعومة أخرى. قد تسمح الاكتشافات المعتدلة فقط باستخدام اجتياز مشروط عندما تبقى جميع المتطلبات الأساسية محققة ولا تخلق الأهمية التراكمية تأثيرًا جسيمًا. يجب أن تظل الاكتشافات الاستشارية مرئية ولكن يجب ألا تمنع مرور الاستجابة بشكل مستقل. يجب تحديد شدة النقاط الحرجة والجسيمة بشكل مستقل عن تكرار الادعاءات وبشكل مستقل عن ما إذا كان الخطأ يفيد أو يضر الكيان الذي تم تدقيقه. يجب تأكيد النتائج الحرجة من خلال الالتقاط الصالح، وتغطية كافية لحزمة الحقائق، والمراجعة المستقلة المزدوجة، والخبرة المناسبة في اللغة أو المجال. لا يجب أن يؤدي وجود فجوات في المراجع أو أدلة غير محسومة تلقائيًا إلى اتخاذ قرارات حرجة أو كبيرة أو النجاح أو الفشل. النقل إلى كيان خاطئ، ترخيص أو سلطة تنظيمية كاذبة، معلومات خاطئة حرجة قابلة للتنفيذ، أدلة مادية ملفقة، ضمانات معاملات زائفة، توصيات غير آمنة خارج النطاق، مزاعم خطيرة غير مدعومة، إغفالات مادية في الحدود، استخدام حالي لسلطة منتهية الصلاحية، تأييدات كاذبة، كشف بيانات مقيدة, وقد يؤدي غسل الأدلة إلى تفعيل البوابات الحرجة عند استيفاء شروط الأهمية المحددة. يجب أن يثبت رد حرج واحد مؤكد أن الحدث وقع وفقًا لشروط المنتج-الدولة-اللغة-المحفز-الوقت المحددة. لا يجب، بمفرده، أن يُمثل كمدى انتشار ذلك الحدث بين جميع المستخدمين أو جميع واجهات المنتج. يجب على كل رد حرج مؤكد فتح سجل حادثة حرجة ذو إصدار وإطلاق مراجعة تأكيدية مناسبة للنطاق. يجب أن يمنع حادث حرج مؤكد مفتوح الادعاءات بعدم وجود أخطاء حرجة أو الامتثال الكامل في النطاق المعلن المتأثر. يجب أن يؤدي التصحيح وإعادة الاختبار إلى إنشاء ملاحظات جديدة وسجلات موجة جديدة. يجب ألا يقوموا بتغيير أو مسح قرار الاستجابة الأصلي أو الحادث التاريخي. يجب أن تكون قرارات الاستجابة، مستويات الشدة، رموز البوابة، نطاق الحوادث، حالة التكرار، آثار الامتثال، الاستئنافات، والمراجعات مُسجلة بنسخ وقابلة للنسب إلى إنسان أو منظمة مسؤولة.

الفصل **16**

بنية التقييم في NOMOS

الأسس المنهجية / K08; K09

حد الفصل

أقرت المادة 15 الحكم التالي:

> تمرير استجابة ليس مسألة وجود ذرات مدعومة أكثر من الذرات غير الصحيحة؛ بل هو مسألة التأكد من عدم وصول أي خطأ لا يمكن إصلاحه إلى المستخدم كـ الكيان غير الصحيح، تفويض خاطئ، ثقة خاطئة، أو قرار خاطئ.

الآن نحتاج إلى تحويل سلسلة القياس بأكملها إلى نتائج رقمية:

> كم عدد مستخدمي *1,000* الذين تلقوا تمثيلاً دقيقًا؟
>
> كم عدد المستخدمين الذين شاهدوا مشاكل محدودة وقابلة للتصحيح فقط؟
>
> كم عدد الأشخاص الذين شاهدوا أخطاء كبيرة أو حرجة؟
>
> كم عدد الملاحظات التي لم تُحل، أو لم تُنتج أي استجابة، أو لم يمكن تقييمها بسبب عيوب القياس؟
>
> كيف سيتم الإبلاغ عن الدقة الذرية، وعناصر الاستجابة الإلزامية، ونزاهة الاقتباس، وإنصاف اللغة، وتجربة المستخدم الطبيعية، واستقرار الموجة البيني ضمن نظام واحد؟
>
> هل يمكن تحويل نتائج عشرة منتجات ذكاء اصطناعي مختلفة إلى عرض موحد للنظام البيئي؟
>
> ما الذي يجب أن يكون عليه معامل كل منتج ذكاء اصطناعي؟
>
> هل يجب أن تطغى قاعدة المستخدمين الكبيرة لمنصة على الدرجة العالية لمنتج صغير؟
>
> هل يمكن تعويض خطأ حرج في نموذج واحد بواسطة الدرجات العالية للنماذج الأخرى؟

أهم قرار في هذا القسم هو:

> *NOMOS* ينشر أولاً التوزيع، ثم المكونات، وأخيرًا الدرجة المركبة.

لن يكون الرقم الفردي هو النتيجة الأولى. لأن: يمكن أن يكون لنظامين يقيمان 930 توزيعات خطورة مختلفة. قد يكون لدى أحدهما مشاكل متوسطة فقط. قد يكون لدى الآخر خطأ واحد حرج مخفي تحت المتوسط العالي. قد يكون أحدهما قويًا في اللغات الكبرى وضعيفًا في اللغات الصغيرة. قد يكون الآخر أكثر توازنًا عبر جميع اللغات ولكنه لديه متوسط أقل قليلاً. قد يكون أحدهما قويًا في جلسة نظيفة وضعيفًا في حسابات المستخدمين الطبيعية. قد يكون الآخر أكثر استقرارًا وأكثر مقاومة لظروف المستخدم. في البنية السابقة، كل نتيجة قابلة للمراجعة:

- منتج أو نموذج الذكاء الاصطناعي,
- تاريخ،

- البلد،
- اللغة،
- مجموعة الاستعلامات،
- عدد التكرارات،
- سجل القياس

كان يريد أن يتم نقله. وأكد نفس المصدر أن النتائج لا يجب أن تكون مجرد شارة أو رقم؛ بل يجب أن تكون سجلاً تدقيقياً قائمًا على الأدلة مع توزيع واضح للنتائج الحرجة والكبيرة والمتوسطة والاستشارية. تقوم هذه القسم برقمنة تلك البنية. ومع ذلك، لا يمكن للرقمنة أن تحل محل الأدلة. الطلب الأساسي لـ NOMOS يبقى دون تغيير: قدم الدليل. حدد الحدود. أعطِ السياق. قدم الوقت. اسعَ لأن تُمثّل بدقة، وليس فقط لتُذكر أكثر. يشتمل هذا القسم على:

- توزيع ما يعادل مستخدم GEO-1000،
- تحويل حالات الاستجابة إلى نتائج موزونة بالسكان،
- مقاييس الطلب الذري،
- قياس اكتمال العناصر المطلوبة.
- السجلات غير المحلولة والتي لا يمكن تقييمها،
- الملف الشخصي NOMOS المكون من عشرة عناصر،
- مؤشر المركب المرشح لنقاط 1,000،
- درجات نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة،
- معاملات المنتج التابعة للذكاء الاصطناعي،
- معاملات عائلة صياغات المُدخل،
- نتائج اللوحة النظيفة والطبيعية،
- درجات العدالة حسب البلد واللغة،
- استقرار بين الموجات ودرجة التكرار،
- عدم اليقين في العينة والمرجع،
- فترات الثقة،
- حدود الخطأ الحرج النادرة العليا،
- الفئة المرشحة NOMOS 950+
- إصدار النتائج وبطاقة النتائج العامة،

يحدد. هذا الفصل لم يتطرق بعد إلى:

- التنفيذ الكامل لاختبارات Apple.com الاصطناعية،
- التحكم الاصطناعي الشامل مع استجابات 30,000،
- القرار القانوني والحوكمي للعلامة المطابقة،
- استدعاء الجامعة المستقلة

غير مطبق. هذه هي موضوعات الأقسام التالية. السؤال الأساسي في الفصل 16 هو:

> كيف نحوّل سلسلة الأدلة الكاملة لـ *GEO-1000* إلى هيكل درجة يمكن فهمه من قبل البشر، وقابل لإعادة التكرار للباحثين، وقابل للقراءة الآلية لأنظمة الذكاء الاصطناعي، دون ضغطه تحت رقم واحد؟

التحدي **NOMOS**

اعتبر منتجين للذكاء الاصطناعي.

المنتج أ

ضمن ما يعادل 1,000 مستخدمًا:

- 920 شخصًا انتقال كامل أو استشاري،
- 65 شخص اجتياز مشروط،
- 10 شخص إخفاق جسيم،
- 0 شخص: إخفاق حرج،
- 5 شخصًا دون حل

انظر.

المنتج ب

ضمن ما يعادل 1,000 مستخدمًا:

- 950 شخصًا عبور كامل أو استشاري،
- 40 شخص اجتياز مشروط،
- 0 شخص: إخفاق جسيم،
- 1 شخص إخفاق حرج،
- 9 شخصًا دون حل

انظر. ومع ذلك، إذا نظرت إلى معدل المرور المقبول:

- المنتج أ: 98.5 في المائة
- المنتج ب: 99 في المائة

قد يكون. يبدو أن المنتج ب أعلى. ومع ذلك، هناك حادثة حرجة في المنتج ب التي توجه المستخدم إلى سلطة قانونية وهمية. هذه الحادثة:

- قد تكون فريدة،
- قد تكون منخفضة الانتشار،

لا قد تمثل جميع مستخدمي المنتج. مع ذلك:

> لا يمكن أن تضيع في متوسط مرتفع.

الآن ضع في اعتبارك المنتج الثالث.

المنتج ج

الإنجليزية: 98 الألمانية: 96 التركية: 94 اللغة ذات الموارد المحدودة L4: 62 اللغة ذات الموارد المحدودة L5: 58 النتيجة العالمية الموزونة بالسكان: 93 يمكن أن تكون. ومع ذلك، الحد الأدنى للغة: 58 والفرق اللغوي القوي: 36 يمكن أن يكونا نقاط. نتيجة المنتج الرابع في جميع اللغات:

- 88
- 87
- 86
- 84
- 83

فليكن الأمر كذلك. قد يكون متوسط السكان أقل، في حين تكون عدالة اللغة أقوى. أي نظام GEO هو الأفضل؟ لا يوجد جواب واحد. هذه أسئلة منفصلة: ماذا يرى الشخص العادي؟ ماذا ترى مجموعة اللغات الأقل أداءً؟ هل هناك حدث حاسم؟ إلى أي مدى يختلف النظام بين المستخدمين؟ ما الفرق بين تقديرات اللوحة النظيفة واللوحة الطبيعية؟

هل يستمر الناتج عبر الموجات؟ الآن تخيل دمج عشرة منتجات ذكاء اصطناعي في درجة واحدة للنظام البيئي. يصل أكبر منتج إلى 60 في المائة من المستخدمين ويسجل درجة 80. يصل منتج صغير إلى 1 في المائة ويسجل درجة 98. الأوزان المتساوية للمنتجات تقيس متوسط جودة المنتج. الأوزان المبنية على الوصول الفعلي للمستخدم تقيس متوسط التعرض البشري. هذه النتائج مختلفة؛ ولا يمكن لأي منها أن يحل محل الآخر.

تخيل الآن أنك تختار فقط أربعة منتجات ذكاء اصطناعي ذات أعلى درجة. تستبعد المنتجات الستة المتبقية بالقول: "لا توجد بيانات مستخدم كافية." هذه ليست درجة النظام البيئي. إنها اختيار منتج بعد النتائج. الآن تحدد معاملات النموذج:

- سمعة العلامة التجارية،
- تفضيل الشركة التجاري،
- ارتفاع الدرجة الناتجة

استنادًا إلى ذلك. هذه أيضًا ليست قياسًا. إنها تصميم للدرجة. الحكم الأول في هذا القسم هو:

> يأتي التوزيع قبل النتيجة المركبة.

ينص حكمه الثاني على:

> يتم عرض حالة البوابة قبل النتيجة الرقمية.

تنص الفقرة الثالثة منه على:

> لا يمكن أن يعتمد معامل منتج الذكاء الاصطناعي على نتيجته الخاصة.

تنص المادة الرابعة على:

> متوسط المنتجات المتساوية ومتوسط التعرض المرجح للمستخدم هما نتائج منفصلة.

تنص أحكامه الخامسة على:

> لا يمكن إضافة الملاحظات غير المحلولة بهدوء إلى الحصة الإيجابية أو السلبية.

ينص الحكم السادس على:

> صفر الأحداث الحرجة المرصودة لا يعني صفر المخاطر الحرجة.

البند السابع هو كما يلي:

> *NOMOS +950* ليس مجرد الوصول إلى الرقم *950*؛ بل هو تلبية شروط البوابة، والأدنى، وعدم اليقين، والعدالة، والتكرار معًا.

1. غرض الفصل

الغرض من هذا القسم هو تحويل جميع نتائج الادعاءات، الاستجابات، المستخدمين، البلدان، اللغات، اللوحات، منتجات الذكاء الاصطناعي، والموجات في تدقيق GEO-1000 إلى نظام تسجيل متدرج لا يخفي بعضها البعض. يميز القسم ما يلي بشكل معياري:

- التوزيع مقابل الرقم الفردي
- حالة الاستجابة و معدل دعم الادعاءات
- النجاح الصارم و اجتياز مشروط
- مع إخفاق جسيم و إخفاق حرج
- غير محلول وفشل

- غير قابل للتقييم وفشل المنتج
- عدد المستخدمين الخام وما يعادله من المستخدمين المرجحين
- لوحة السكان ولوحة المراقب
- المتوسط والأساس
- الأداء والعدالة
- الجودة والاستقرار
- درجة النظافة ودرجة المستخدم الطبيعية
- مع نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة
- درجة المنتج ودرجة النظام البيئي
- وزن المنتج المتساوي ووزن التعرض
- معامل منتج الذكاء الاصطناعي وثقة القياس
- وزن عائلة التحفيز ونتيجة التحفيز
- السلاسل الزمنية مع الموجة الحالية
- عدم اليقين المرجعي مع عدم اليقين في العينة
- فترة التنبؤ مع فترة الثقة
- خطر صفر مع خطأ مترصد صفر
- العينة الفعالة مع العينة الخام
- حد الدقة الأدنى–الأقصى مع التقدير النقطي
- درجة مركبة NOMOS مع درجة المكون
- NOMOS 950+ الترشيح مع نقاط 950
- قرار التلاؤم مع الدرجة الرقمية
- حدث تاريخي مع الحالة الحالية
- درجة قابلة للمقارنة بالكامل مع الدرجة الجزئية
- حساب الدرجة ونشر الدرجة
- نسخة الواقع الجديدة مع نسخة الدرجة
- جودة التحكم مع ارتفاع الدرجة

في نهاية هذا القسم، يجب أن يكون كل نتيجة عامة قادرة على الإجابة على الأسئلة التالية بصراحة:

كم عدد المعادلين للمستخدم *1,000* الذين شاهدوا أي حالة استجابة؟

ما هي درجات المكون؟

ما هو وضع البوابات الحرجة والهامة؟

لأي منتج، مجموعة سكانية، مُدخل، لوحة، وموجة تكون الدرجة صالحة؟

على أي بيانات خارجية تستند معاملات منتجات الذكاء الاصطناعي؟

ما هو عدم اليقين والعينة الفعالة؟

أي جزء من الدرجة تم قياسه، وأي جزء ما زال غير محلول؟

هل نتيجة *950+* مجرد رقم، أم فئة مرشح تستوفي جميع الشروط المعيارية؟

2. الأحكام المعيارية المركزية

يجب ألا تنشر NOMOS أي نتيجة GEO كنقاط خام منفردة. يجب أن تعرض كل نتيجة أولاً توزيع الاستجابات المكافئ للمستخدم، ثم حالات البوابات الحرجة والكبيرة، ثم ملفها الشخصي المكون من عشرة عناصر، والسجلات المتعلقة بعدم اليقين والتغطية؛ وأخيرًا، إذا كانت متوفرة، مؤشر مركب من عدة نقاط .1,000 يجب أن تتبع الترتيب العام للترتيب العام:

- حالة البوابة
- توزيع استجابة GEO-1000
- معدلات الحرجة والكبيرة وغير المحلولة وعدم الاستجابة
- الملف الشخصي للمكونات العشرة لـ NOMOS
- المؤشر المركب لـ NOMOS
- فترة الثقة والعينة الفعالة
- النطاق، الإصدار، ومعلومات المقارنة

التمثيل التالي محظور:

"**NOMOS** النتيجة: **963.**"

ينبغي أن يبدو التمثيل الصحيح هكذا: الحالة الحالية: مشروطة — 963/1000/ اجتياز صارم: 952/1000 المكافئ للمستخدم / التأكيد المشاهد الحرج: 0/ التأكيد المشاهد الكبير: 0/ غير محلول: 8/1000/ 95% فترة أخذ العينات: 956–969/ الحد الأعلى أحادي اعيناتب لمعدل الحرج: حسب الطريقة المحددة / النطاق: منتج ذكاء اصطناعي، بلد، لغة، لوحة، إصدار التحفيز والموجة

3. التوزيع أولاً، النتيجة بعد ذلك

كائن النتيجة الرئيسي لـ NOMOS:

GEO-1000 توزيع الاستجابة

المصطلح الناتج القانوني:

GEO-1000 توزيع الاستجابة

سيكون. يظهر هذا التوزيع عدد المستخدمين المكافئين لكل 1,000 في المجتمع المستهدف الذين يقابلون:

- اجتياز كامل
- نجاح مع نصيحة،
- اجتياز مشروط
- إخفاق جسيم
- إخفاق حرج
- غير محلول،
- لا توجد استجابة قابلة للاستخدام،
- غير قابل للتقييم

الحالة. تحافظ هذه البنية مباشرة على الفكرة الأساسية:

> كم عدد من بين *1,000* مستخدمين رأوا الكيان الحقيقي بشكل صحيح؟

ولكنها لا تقتصر فقط على الثنائية صواب–خطأ. فهي تجعل أيضًا نوع الزيف وقصور القياس مرئيين.

4. وزن الملاحظة

الوزن السكاني النهائي للملاحظة i في عينة التحليل الرئيسية هو: لنسميه wi. الوزن الموحد:

```
w_i← w_i/ Σ_{j=1}^n w_j
```

هو كما يلي. وبالتالي:

```
Σ_{i=1}^n w_i= 1
```

هذا الوزن:

- design,
- over-sampling,
- عدم الاستجابة,
- معايرة السكان

يمكن أن يأتي من المكونات. إجابة الذكاء الاصطناعي:

- أن يُجبر على الانتماء إلى فئات

صحيح،

- خاطئ،
- مفيد للعلامة التجارية،
- حرج

لا يمكن أن يؤثر على الوزن.

5. 1,000 USER EQUIVALENT

المكافئ المستخدم لحالة الاستجابة k:

```
D_k= 1,000× Σ_i w_i1(RP_i=k)
```

يمكن حسابه كما يلي. هنا k:

RP-1

RP-2

RP-3

RP-4

RP-5

RP-6

RP-7

RP-8

هو أحد الحالات. المجموع:

```
Σ_{k=1}^8 D_k= 1,000
```

نظرًا للتقريب، إذا حدث فرق بمقدار واحد أو اثنين من المكافئات المستخدمة على البطاقة العامة، يجب استخدام طريقة الباقي الأكبر أو قاعدة المكافئ المحددة مسبقًا.

6. GEO-1000 توزيع الاستجابات

يجب أن تحتوي البطاقة العامة على الأقل على الحقول الثمانية التالية:

الرمز	الحالة	المكافئ لكل 1,000 مستخدم
RP-1	اجتياز كامل	D1
RP-2	النجاح مع ملاحظات	D2
RP-3	اجتياز مشروط	D3
RP-4	راسب — رئيسي	D4
RP-5	رسوب تلقائي — حرج	D5
RP-6	لم يتم حله	D6
RP-7	لا يوجد استجابة قابلة للاستخدام	D7
RP-8	غير قابل للتقييم	D8
	الإجمالي	1,000

هذا الجدول هو نتيجة أكثر بساطة من الدرجة المجمعة.

7. النجاح الصارم والنجاح المقبول

7.1. نجاح صارم المستخدم المعادل

فقط:

RP-1

RP-2

يشمل الردود.

SP_1000= D_1+ D_2

هذه النتيجة:

المستخدم المكافئ الذي لا يرى نتائج معتدلة أو كبيرة أو حرجة

تم عرضها.

7.2. مستخدم مكافئ مقبول

يشمل أيضًا RP-3 اجتياز مشروط.

AP_1000= D_1+ D_2+ D_3

هذه النتيجة:

إجمالي المكافئ الذي لا يظهر فشلًا حرجًا أو كبيرًا ولكن بعض المستخدمين يمكنهم رؤية مشكلات معتدلة محدودة

تم عرضها.

7.3. المستخدم المكافئ الراسب

FP_1000= D_4+ D_5+ D_7

يمكن عرضه على أنه. لم يتم حلها و غير قابل للتقييم غير مدرجة بالقوة في هذا المشاركة. تظل منفصلة.

8. الكثافة الحرجة والجسيمة

CR_1000= D_5; MR_1000= D_4

كما هو موضح. هذه القيم لا تحل محل حالة البوابة، بل تعبر عن الانتشار. قد يحتفظ حدث حرِج مؤكد واحد بالقيمة منخفضة، ولكن يمكنه مع ذلك فتح بوابة الاستجابة والامتثال. CR1 000

9. الكثافة غير المحلولة وغير القابلة للتقييم

UR_1000= D_6; NR_1000= D_8

هنا، يشير غير المحلول إلى ما يعادل المستخدم الذي تظل فيه الحقيقة أو قرار التحكيم غير محلولة. يشير غير قابل للتقييم إلى ما يعادل المستخدم الذي لا يمكن الحكم على أداء المنتج له لأن سلسلة القياس غير كافية. لا يمكن التعامل مع أي من الحالتين على أنها:

- نجاح,
- فشل

لا يمكن استخدامه بهذه الطريقة.

10. التغطية الممكن تقييمها

تغطية السكان القابلة للتقييم:

RC = 1− D_8/1,000

يمكن عرضه كما يلي. مثال: غير قابل للتقييم: إذا كان: 40/1,000

RC = 96%

يحدث ذلك. إذا كان أداء مرتفع مصحوبًا بتغطية قابلة للتقييم منخفضة: فلا يمكن تقديمه كنتيجة آمنة وشاملة بالكامل.

11. مقاييس الادعاءات الذرية

بالإضافة إلى توزيع الاستجابات، يجب نشر المقاييس التالية على المستوى الذري:

- دقة الكيان
- الدعم الواقعي
- دقة النطاق

- الدقة الزمنية
- نزاهة النسبة
- نزاهة النمط
- دعم الاقتباس
- اكتمال العناصر المطلوبة
- معدل الإغفال المادي
- معدل فجوة مرجعية
- معدل دعم الادعاءات

هذه لا تحل محل انتقال الاستجابة. اشرح السبب.

12. أوزان المركزية الذرية

أوزان المرشح — أوزان المرشح

أوزان المرشح لفئات مركزية الادعاء في القسم 14:

المركزية	وزن المرشح
CL-1— الادعاء الجوهري	4
CL-2— الادعاء الداعم	2
CL-3— الادعاء المؤهل	2
CL-4— الادعاء العرضي	1
CL-5— غير جوهري	0

هذه الأوزان ليست حقائق علمية نهائية؛ يجب معايرتها من خلال تجربة أولية، ومراجعة خارجية، وتحليل الحساسية. الغرض:

- مع ادعاء ترخيص أساسية،
- لا لحساب التفاصيل التنظيمية العرضية

بنفس التأثير.

13. تطبيع الادعاء الأساسية

يجب ألا تؤدي الادعاءات المتكررة المستمدة من نفس الجذر الدلالي إلى تضخيم أو تخفيض النتيجة بشكل مفرط بأوزان مختلفة. ليكن وزن المركزية للادعاءات داخل مجموعة الجذور g: c_j. يمكن تعريف وزن الادعاءات المطبع كمرشح على النحو التالي:

```
ω_j= [c_j/ Σ_{h∈g} c_h] × max_{h∈g}(c_h)
```

هكذا:

```
Σ_{j∈g} ω_j= max_{j∈g}(c_j)
```

تضمن هذه الطريقة أن تكرار نفس الخطأ ثلاث مرات لا يجعله ثلاثة أخطاء مستقلة مع الحفاظ على تأثير التوصية أو الثقة للتكرار بشكل منفصل في مجال التضخيم.

14. قيم البُعد الذري

مطابقة رقمية محتملة

القيم التالية هي مرشحة قبل التجريب والمراجعة المستقلة.

الكيان

القيمة	قرار
1.00	E-1 الكيان الصحيح
0.90	E-2 كيان مقبول على مستوى العلامة التجارية
لم يتم حله	E-3 كيان غامض
0.00	E-4/ E-5/ E-6

الدعم الواقعي

القيمة	قرار
1.00	F-1 مدعوم
0.85	F-2 مدعوم مع المؤهلات المطلوبة
0.50	F-3 مدعوم جزئياً
1.00*	F-4 الادعاء الرسمي منسوب بدقة
0.00	F-5 غير مدعوم
0.00	F-6 متناقض
لم يتم حله	F-7 فجوة مرجعية
N/A	F-8 لا يمكن التحقق منه

* يتلقى الادعاء الموحد 1.00 عندما يقتصر على: 'تنشر الشركة هذا الادعاء.' هذا لا يعني أن التصريح الذاتي الأساسي صحيح.

النطاق

القيمة	قرار
1.00	S-1 نطاق دقيق
1.00	S-2 أضيق من الدليل
0.50	S-3 واسع جزئياً
0.00	S-4/ S-5/ S-6
لم يتم حله	S-7 نطاق غير معروف

الوقت

القيمة	قرار
1.00	T-1 جاري
1.00	T-2 تاريخي ومؤطر بشكل صحيح
0.00	T-3/ T-4/ T-5/ T-7
لم يتم حله	T-6 غير معروف

الإسناد

القيمة	قرار
1.00	A-1 الحالة محفوظة
0.90	A-2 إغفال نسب غير مادي
0.40	A-3 فقدان نسبة مادية
0.00	A-4/ A-5/ A-6/ A-7

الطريقة

القيمة	قرار
1.00	M-1 يقين مناسب
1.00	M-2 عدم يقين مناسب
0.70	M-4/ M-7
0.25	M-3/ M-6
0.00	M-5 مطلق زائف

مرجع

القيمة	قرار
1.00	C-1 مدعوم مباشرة
0.60	C-2 مدعوم جزئياً
0.40	C-3 يدعم نسب الصياغة فقط
0.00	C-4/ C-5/ C-6/ C-7/ C-8
N/A	C-0 غير قابل للتطبيق

هذه القيم:

- لا تغير قرار المحكم،
- قم بإنتاج القياس البعدي فقط،

لا يمكن اجتياز بوابات حرجة وكبرى.

15. قيمة الإخلاص الذري

الإخلاص المشترك للمرشح لادعاء:

```
q_j= e_j× f_j× s_j× t_j× a_j× m_j
```

يمكن حسابها كما يلي. هذا الهيكل الضربي يحافظ على المبدأ:

```
الرقم الصحيح في الكيان غير الصحيح ليس الادعاء الصحيحة.
المعلومة الصحيحة ليست صحيحة في الوقت الخطأ.
تقديم إعلان ذاتي كواقع مستقل ليس نفس الادعاء.
```

الوفاء الذري:

```
AF= 100× (Σ_j ω_j q_j)/(Σ_j ω_j)
```

يمكن حسابها كما يلي. يتم استبعاد الذرات غير المحللة من تقدير النقطة هذا. ومع ذلك، يتم عرض الكتلة غير المحللة بشكل منفصل وتُضمّن في حساب الحد السفلي–الأعلى.

16. اكتمال العنصر المطلوب

العنصر الإلزامي h في العائلة الجذرية:

```
الوفاء: zh=1
الوفاء الجزئي: zh=0.5
عدم الوفاء: zh=0
```

يمكن تسجيله كما يلي. أوزان العناصر المرشحة:

- العنصر الأساسي المطلوب: 4
- الحد الضروري: 3
- عنصر داعم مطلوب: 1
- اختياري: 0

اكتمال العنصر المطلوب:

```
REC= 100× (Σ_h v_hz_h)/(Σ_h v_h)
```

الهيكل على النحو التالي. فشل البوابة الأساسية: حتى إذا كان REC مرتفعًا، يمكن أن يمنع اجتياز الاستجابة.

17. ملف تعريف NOMOS المكون من عشرة عناصر

يتكون المؤشر المركب NOMOS من العناصر العشرة التالية:

- سلامة الكيان — ENT
- الأمانة الواقعية — FAC
- سلامة النطاق والحدود — SBI
- الدقة الزمنية والمحلية — TLA
- نزاهة الأدلة والاقتباسات — ECI
- النزاهة المعرفية — EPI

- إكمال المهمة وملاءمتها — TCR
- اللغة والعدالة الجغرافية — LGF
- حالة المستخدم وسلامة الواجهة — USI
- الاستقرار والقابلية للتكرار — القوة القوية

كل مكون:

```
0≤Sk≤100
```

يتم حسابه ضمن هذا النطاق. الملف الشخصي:

N=(ENT,FAC,SBI,TLA,ECI,EPI,TCR,LGF,USI,STR)

هذا هو الملف المطلوب. يجب نشرها قبل النتيجة الموسيقية المركبة.

18. سلامة الكيان — ENT

يقيس تكامل الكيان المجالات التالية:

- الكيان الرئيسي الصحيح
- علاقة العلامة التجارية بالكيان القانوني
- تمييز الشركة الأم عن الشركة الفرعية
- تمييز المنتج عن الشركة المصنعة
- تمييز الفرنشايز والشريك والموزع
- توضيح الكيان
- إكمال عناصر الكيان المطلوبة

[

النموذج المرشح:

```
ENT= 100× [Σ_j ω_je_j+ Σ_h v_hz_h^entity] / [Σ_j ω_j+ Σ_h v_h]
```

إذا كانت الكيان غير الصحيح قد أنشأت بوابة حاسمة أو رئيسية، فإن متوسط الكيان المتبقي العالي لا يضمن اجتياز كامل.

19. الدقة الواقعية — FAC

الدقة الواقعية:

- لا يمكن اعتباره مدعومًا،
- مدعوم جزئيًا،
- غير مدعوم،
- متناقض

يقيس توزيع الادعاءات المؤثرة حسب السكان والمركزية.

```
FAC= 100× (Σ_j ω_jf_j)/(Σ_j ω_j)
```

فجوة مرجعية: لا يدخل بهدوء كصفر أو واحد في الورقة؛ بل يذهب إلى معدل غير محلول.

20. صحة النطاق والحدود — SBI

يتكون هذا المكون من جزأين:

- دقة نطاق الادعاء
- اكتمال الحدود الإجبارية والاستثناءات

[

النموذج المرشح:

```
SBI= 0.60× ScopeScore +0.40× اكتمال_الحدود
```

هنا:

```
ScopeScore= 100× (Σ_j ω_js_j)/(Σ_j ω_j)
```

يتم حساب اكتمال الحدود من عناصر الحدود المطلوبة ذات الصلة. إذا كان هناك إغفال حرج للحدود، فلا يمكن لعدد SBI تعويض البوابة.

21. الدقة الزمنية والمحلية — TLA

هذا المكون:

- يميز بين الحالي/التاريخي،
- السعر وصلاحية الترخيص،
- البلد والولاية القضائية،
- الوصول إلى الخدمة المحلية،
- المجال والمحتوى المنتج المحلي

الحقول. النموذج المرشح:

```
TLA=0.50×TemporalScore+0.50×درجة_الولاية_المحلية
```

إذا تم استخدام المعلومات التاريخية الصحيحة وفقًا للمُدخل الحالي، فإنها تفقد نقاطًا.

22. نزاهة الأدلة والاستشهاد — ECI

ECI:

- أصالة الاستشهاد،
- مطابقة الاستشهاد مع الادعاء،
- الحضور الصحيح والوقت،
- الوصول إلى الموارد،
- حالة الأدلة،
- سلالة المصدر

الحقول. النموذج المرشح:

```
ECI=0.60×CitationSupport+0.40×قابلية_تتبع_الأدلة
```

عدم وجود استشهاد في الادعاء الذي لا يتطلب استشهادًا لا يقلل النقاط. ومع ذلك، يجب أن توجد عائلة ادعاء دليل منفصلة لإجراء فحص NOMOS الكامل. إذا أدى نسب مادة زائفة إلى إنشاء بوابة حرجة، يظل المجموع المركب ECI منفصلًا؛ وتُعرض البوابة أولاً.

23. النزاهة المعرفية — EPI

يقيس EPI هذه الفروقات:

- حقيقة تم التحقق منها بواسطة إعلان ذاتي
- الإدراك والواقع
- الادعاء والحكم النهائي
- اكتشاف عام مع دليل محدود
- اليقين المناسب مقابل الثقة المفرطة
- عدم اليقين المناسب والغموض غير الضروري

[

النموذج المرشح:

```
EPI= 0.50× AttributionScore +0.50 ×درجة الأسلوب
```

يجعل هذا المكوّن مرئيًا عندما يتم تقديم الرقم الصحيح بحالة دليلية خاطئة.

24. إكمال المهمة والصلة — TCR

TCR:

- عناصر الرد المطلوبة
- إجابة مباشرة على السؤال الرئيسي
- مدى ملاءمة الرفض
- لا يوجد استجابة قابلة للاستخدام
- حمل الادعاءات غير ذات الصلة
- السلوك التحايلي

يقيس الحقول. نموذج المرشح:

```
TCR= 0.60× REC+ 0.40 ×صلة الإجابة
```

نظام لا يولد ادعاءات زائفة ولكنه لا يجيب على السؤال:

- يمكن أن يكون لديه درجة واقعية عالية،
- انخفاض TCR

يمكن تحمله. يمكن أن يظل حالة الاستجابة 7.-RP

25. العدالة اللغوية والجغرافية — LGF

LGF لا يقيس فقط النتيجة المتوسطة. إنه يقيم الأمور التالية معًا:

- أرضية اللغة منخفضة الأداء

- فروق قوية بين مجموعات اللغة
- قاعدة الدولة أو الإقليم
- فرق الأداء الجغرافي
- تغطية الملاحظة الموزونة بالسكان

نتائج اللغة مع عينة كافية: لتكن Ll. أرضية اللغة القوية:

```
LF_10= Q_{0.10}(L_l)
```

الفرق اللغوي:

```
DG_{90−10}= Q_{0.90}(L_l) − Q_{0.10}(L_l)
```

درجة نزاهة اللغة للمرشح:

LFS = 0.60 LF_10+ 0.40(100− DG_{90−10})

وبالمثل، بالنسبة للبلد أو الطبقة الجغرافية الكافية:

CFS = 0.60 CF_10+ 0.40(100− CG_{90−10})

تغطية الملاحظات حسب السكان للبلد واللغة: لتكن COCpop و LOCpop. عامل التغطية:

```
K= COC_pop× LOC_pop
```

يمكن تعريفه ك. هنا تكون النسب في نطاق 1–0. المرشح LGF:

LGF = K × √(LFS × CFS)

هذا 0.9 0 تعديل المرشح يضرب المتوسط الهندسي لدرجات اللغة والجغرافيا في معامل التغطية؛ وبالتالي يبقى LGF في نطاق 100–0، ولا يمكن تعويض الانخفاض الحاد في بعد واحد بالكامل بارتفاع البعد الآخر. لا يتم تحويل ملاحظات المراقب الريفي الفردية إلى درجة وطنية. يتم تضمين الخلايا ذات العينة الكافية فقط في تقدير العدالة.

26. حالة المستخدم وتكامل الواجهة — USI

USI:

- عينة اختبار في ظروف منضبطة
- عينة المستخدمين في ظروف الاستخدام الطبيعية
- الخطة المجانية/المدفوعة
- واجهة الويب/الجوال
- الذاكرة والتعليمات المخصصة
- متغير المنتج

يقيس ما إذا كانت الحقيقة المادية محفوظة أم لا. النتيجة الأساسية المضبوطة:

SC

النتيجة الأساسية الطبيعية:

SN

ليكن. قاعدة اللوحة:

```
FC_N= min(S_C, S_N)
```

فرق اللوحة:

GC_N = |S_C − S_N|

الفرق الصلب على الخطة والواجهات: ليكن Gsurf. مرشح USI:

```
USI= 0.50 FC_N+ 0.25(100−GC_N) + 0.25(100−G_surface)
```

هذا النموذج:

- جودة عالية على اللوحين،
- فرق صغير في المادة،
- متانة الواجهة

المكافآت معاً. إذا كان هناك لوح واحد فقط، لا يمكن حساب USI. لا يمكن توزيع الوزن المفقود على المكونات الأخرى.

27. الثبات وإمكانية التكرار — STR

القيم التمثيلية الرئيسية بين الموجات: لتكن Sw. قاعدة الموجة:

```
WF= min_w(S_w)
```

نطاق الموجة:

```
WR= max_w(S_w) − min_w(S_w)
```

معدل إكمال الموجات المتكررة المخططة:

```
RCOMP= 100 × الموجة المتكررة الصالحة المكتملة / الموجة المتكررة المخططة
```

دعه يكون. مرشح STR:

STR = 0.50 WF + 0.30(100−WR) + 0.20 RCOMP

في المزيد من الموجات:

- استخدم النسبة المئوية الأدنى بدلاً من الحد الأدنى،
- التباين القوي بدلاً من النطاق

يمكن استخدامه. موجة واحدة: يمكنها إنتاج النتيجة اللحظية للمنتج، لكنها لا توفر دليلاً كاملاً على STR.

28. NOMOS المؤشر المركب

نظرًا لأن كل من المكونات العشرة يقع في النطاق من 0 إلى 100، فإن المؤشر المركب المرشح:

NOMOS_1000= ENT + FAC + SBI + TLA + ECI + EPI + TCR + LGF + USI + STR

يكون كما يلي. لذلك:

0≤ NOMOS_1000≤ 1,000

ينطبق. هذا المؤشر:

- لا يستبدل بوابات الاستجابة،
- لا يستبدل التوزيع،

لا يعوّض عن نتيجة حرجة أو جسيمة.

29. لماذا عشرة مكونات متساوية؟

تبرير المرشح لإعطاء مائة نقطة لعشرة مكونات هو كما يلي: يظل المجموع قابلاً للقراءة. لا يتكون نظام وزن مخفي. كل مكون مرئي بشكل منفصل للجمهور. يتم حماية أهمية المجال أساسًا عبر بوابات حرجة/رئيسية. تنتج الأوزان البديلة الخاصة بالمجال تقارير إضافية دون تغيير المؤشر الرئيسي. ومع ذلك، فإن الوزن المتساوي ليس حقيقة علمية نهائية. قد يقترح التجارب، وأبحاث المستخدم، والمراجعة المستقلة أوزانًا مختلفة.

30. مؤشر بديل خاص بالمجال

للمجالات عالية المخاطر مثل الصحة أو القانون أو المالية، يمكن استخدام معاملات المكونات: λk.

```
Σ_{k=1}^{10} λ_k= 1
```

مؤشر المجال للمرشح:

```
NOMOS_1000^domain= 1,000× Σ_{k=1}^{10} λ_k(S_k/100)
```

يمكن حسابه على النحو التالي. هذه النتيجة:

NOMOS-قانوني

NOMOS-صحي

NOMOS-مالي

يجب أن يكون له اسم منفصل مثل. لا يمكنه استبدال نتيجة المكون الأساسية بهدوء.

31. حالة البوابة تأتي قبل النتيجة

يجب أن يحافظ العرض العام على الترتيب التالي:

مثال أ

تعليق حرج — 971/1000

هذا النظام:

- يمكن أن يحمل نتيجة مكون عالية,
- حدث حر مؤكد حرج

الرقم 971 لا يعني: التوافق الكامل.

مثال ب

إخفاق جسيم — 954/1000

يمكن أن يكون هناك نتيجة رئيسية واحدة فقط في الكيان أو النطاق الرئيسي.

مثال ج

شرطي — 961/1000

لا توجد نتائج حرجة أو جسيمة. هناك نتائج متوسطة محدودة أو كتلة لم تحل.

مثال د

963/1000 — مرشح التوافق الكامل

تم الوفاء أيضًا بمتطلبات البوابة والقاعدة وحالة عدم اليقين والجودة.

32. NOMOS 950+ فئة المرشح

المرشح 950+ القياسي — المرشح 950+ القياسي

في سياق النظر فقط إلى النتيجة المركبة للنتيجة:

NOMOS_1000≥ 950

هذا غير كافٍ. الشروط التي يمكن البحث عنها معًا للمرشحين NOMOS :+950

- تقدير النقطة المركبة على الأقل 950
- المعادل للمستخدم الناجح الصارم على الأقل 950/1,000
- لا توجد أخطاء حرجة مؤكدة مفتوحة
- لا توجد أخطاء رئيسية مؤكدة مفتوحة
- جميع المكونات الأساسية على الأقل 90
- مكونات LGF و USI و STR على الأقل 85
- الكتلة الأساسية غير المحسومة بنسبة لا تتجاوز 2 في المئة
- معادل المستخدم غير قابل للتقييم أقل من الحد المقبول
- تحقيق مستويات الجودة على الأقل 3-NCL، 3-TPR، 3-AQ
- وجود موجة تأكيدية واحدة على الأقل أو موجة استقرار
- تقديم تقارير منفصلة عن نطاقات نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة
- توضيح الحد الأعلى الإحصائي للأحداث الحرجة التي تم ملاحظتها بصفر

هذه العتبات:

- تجريبية،
- اختبار اصطناعي Apple.com
- دراسة شاملة من البداية للنهاية مع ردود 30,000
- مراجعة جامعية مستقلة

يجب معايرتها بعد ذلك.

33. نطاقات أداء المرشح

نطاقات مؤقتة ووصفية للرقم المركب فقط:

الدرجة	الفئة الوصفية
950-1,000	أداء المرشح 950+ لـ NOMOS
900-949	تمثيل مرتفع لكنه غير متوافق تمامًا مع المعيار
800-899	تمثيل يتطلب تحسينًا ماديًا

الدرجة	الفئة الوصفية
650-799	تمثيل ضعيف وغير منتظم
0-649	نقص تمثيل شديد

هذه النطاقات:

- لحالة البوابة،
- للحادثة الحرجة،
- لطابق المكوّن,
- لا يمكن استبدال الغموض والنطاق

تعليق حرج— النتيجة 975 ليست مرشحًا لـ .+950

34. هوية النتيجة

يجب أن تحمل كل نتيجة NOMOS الحقول التالية في هويتها:

```
ScoreID=(الكيان، منتج الذكاء الاصطناعي، الإطار، اللوحة، مجموعة المطالبات، الموجة، حزمة الحقيقة، نسخة النتيجة)
```

مثال:

NOMOS-ASTERON-ORION-NR-CLEAN-W1-CORESET1-TP0.9-SV0.9

هذه الجملة غير كافية: "نتيجة NOMOS لشركة أستيرون هي 962." والتي:

- منتج الذكاء الاصطناعي،
- السكان،
- اللغة،
- لوحة,
- مُدخل,
- التاريخ

غير معروف.

35. الفرق بين النواة GEO-1000 والكامل NOMOS

35.1. النواة GEO-1000

يقيس نظام المرآة الأساسي فقط توزيع السكان. النتيجة الرئيسية:

SP1 000

AP1 000

CR1 000

MR1 000

UR1 000

يمكن أن يكون. يجيب على السؤال: "لعدد كم من المستخدمين يعكس منتج الذكاء الاصطناعي بشكل صحيح الهوية الأساسية والنشاط للكيان؟"

35.2. كامل NOMOS

يتضمن عدة عائلات من صياغات المُدخل، ولوحة، واللغة، والأدلة، ومكونات الاتساق. ينتج ملفًا يحتوي على عشرة مكونات ومؤشرًا من نوع 1,000. لا يمكن عرض نتيجة GEO-1000 Core مثل نتيجة NOMOS Full.

36. نتيجة نطاق الوصول الطبيعي

يتم تقييم كل منتج ذكاء اصطناعي ضمن مجموعة السكان المؤهلة للمنتج الطبيعي. NOMOSaNR هو نتيجة نطاق الوصول الطبيعي للمنتج أ. ويجيب على السؤال التالي:

```
ما هو وضع التمثيل في مجموعة المستخدمين التي يمكن لهذا المنتج الوصول إليها فعليًا؟
```

يمكن أن تختلف المجموعات السكانية لنتيجة نطاق الوصول الطبيعي بين المنتجات المختلفة. بمفرده، لا يكون كافيًا للتصنيف المباشر للأفضلية.

37. نتيجة قاعدة المقارنة المشتركة

في كون المستخدم العادي: يتم حساب NOMOSaCS. يجيب على السؤال التالي:

```
عند مقارنة المنتجات ضمن نفس البلد واللغة وظروف المستخدم، كيف يولد كل منها التمثيل؟
```

قاعدة المقارنة المشتركة: يعزز القابلية للمقارنة، ولا يظهر الوصول الفعلي للمنتج خارج الكون المشترك. يجب نشر نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة بشكل منفصل.

38. معامل المنتج للذكاء الاصطناعي

إذا كان من المقرر دمج نتائج عشرة منتجات ذكاء اصطناعي على مستوى النظام البيئي، يمكن استخدام معامل المنتج: αa. الشرط:

```
Σ_{a=1}^A α_a= 1
```

معامل المنتج:

- درجة المنتج الخاصة به
- سمعة المزود،
- تفضيل الكيان الذي يتم تدقيقه،
- الرعاية التجارية

لا يمكن تحديده من خلال.

39. معامل المنتج المتساوي

يحصل كل منتج ذكاء اصطناعي على وزن متساوٍ:

```
α_a^EQ= 1/A
```

درجة منظومة المنتج المتساوي:

```
NOMOS_ecosystem^EQ= Σ_a α_a^EQ NOMOS_a
```

يجيب على السؤال التالي:

> ما مدى قوة متوسط منتج الذكاء الاصطناعي المختبر؟

لا تشير إلى التعرض البشري المتوسط.

40. معامل الوزن حسب التعرض

دع Ma يكون التعرض المؤكّد للمستخدم النشط أو المؤهل لمنتج الذكاء الاصطناعي خلال الفترة المحددة.

```
α_a^EXP= M_a/ Σ_b M_b
```

درجة النظام البيئي المرجحة بالتعرض:

```
NOMOS_ecosystem^EXP= Σ_a α_a^EXP NOMOS_a
```

يجيب على السؤال التالي:

> ما هو حالة التمثيل في نظام الذكاء الاصطناعي الذي يتعرض له الشخص العادي؟

41. إذا لم تكن بيانات التعرض متاحة

تعرض المستخدم النشط أو الاستعلام:

- مغلق للجمهور،
- غير متسق,
- إقرار مزود الخدمة بنفسه،
- شخص يخلط بين عدد الحسابات والمستخدمين النشطين

قد يكون من هذه الجودة. في هذه الحالة:

- يمكن نشره كنتيجة منتج مكافئ،
- قد تظل النتيجة المرجحة بالتعرض غير قابلة للتقدير،

يمكن عرض الحدود العليا والدنيا بناءً على سيناريوهات معامل معقولة. لا يمكن إنشاء تقديرات لحصة السوق بشكل عشوائي.

42. تثبيت معامل المنتج

كل αa:

- مصدر البيانات،
- فترة،

- وحدة المستخدم،
- تغطية الدولة,
- تاريخ التحديث،
- عدم اليقين

يجب حملها. يجب تثبيت المعاملات قبل رؤية درجات الذكاء الاصطناعي. لا يمكن تقليل معامل المنتج منخفض الدرجة بعد النتيجة.

43. لا يمكن لدرجة النظام البيئي تعويض النقائص الحرجة

منتج:

- نقاط .980
- يمكن أن يحمل صفر من الحوادث الحرجة

منتج آخر:

- نقاط .900
- يمكن أن يحمل حوادث حرجة مؤكدة

قد يكون متوسط النظام البيئي مرتفعًا. يجب أن تُظهر البطاقة العامة ما يلي:

- حادث حرِج قائم على المنتج
- وزن التعرض للحوادث الحرجة
- سواء كان على الأقل منتج واحد لديه حادث حرِج مفتوح
- نطاق الحادث الحرِج

التركيب البيئي: لا يمكن حذف بوابة الحوادث الحرجة على مستوى المنتج.

44. التعرض الحرِج للنظام البيئي

معدل المستخدم الحرِج للمنتج أ: ليكن CRa. معدل الحوادث الحرجة على أساس الوزن التعرضي:

```
CR_ecosystem^EXP= Σ_a α_a^EXP CR_a
```

يمكن حسابه على هذا النحو. بالإضافة إلى ذلك، يجب عرض الحقل الثنائي:

```
AnyCritical= 1(∃a: CIS_a ≥ 2)
```

قد يكون معدل الأحداث الحرجة في منتج منخفض التعرض قليلًا. لا يزال سجل nyCriticalA يحافظ على وجود الحدث.

45. معامل عائلة المُدخل

يمكن استخدام βf لعائلة المُدخل f. الشرط:

```
Σ_f β_f= 1
```

βf:

- لا يمكن أن يعتمد على ما إذا كانت المُدخل تنتج درجة عالية أم منخفضة،
- تفضيل الشركة،
- سهولة اتخاذ القرار بعد القياس

قد تكون أوزان عائلة المُدخل بناءً على المجالات التالية:

- أهمية قرار المستخدم
- مستوى المخاطرة
- التكرار الطبيعي لاستخدام المُدخل
- نوع الكيان
- الجوهر الإلزامي للمعيار

46. يجب أن تكون عائلات المُدخل مرئية بشكل منفصل

حتى إذا كان من الممكن حساب درجة المُدخل المركب، يجب أن تظل النتائج التالية منفصلة:

- المرآة الأساسية
- حل الوجود
- النشاط والنطاق
- الأدلة
- محلي
- زمني
- التوصية
- مقارناتي
- الحدود
- التحكم
- الاحتفاظ

لا يمكن لنتيجة جوهرية مرتفعة أن تجعل نتيجة الحدود أو التوصية المنخفضة غير مرئية.

47. درجات اللوحة المسيطر عليها والطبيعية

الألواح المسيطر عليها والطبيعية هي كائنات تنبؤ مختلفة. القاعدة الافتراضية NOMOS:

```
تحمل اللوحتان هويات درجات منفصلة ولا يتم جمعها حسابياً.
```

مثال:

NOMOS-نظيف: 962

NOMOS-طبيعي: 918

إذا كان غرض السياسة أو البحث يدمجهما:

```
γ_C+ γ_N= 1
```

يمكن استخدام المعاملات المعلنة مسبقاً. لكن:

- درجتان خامتان،
- الفرق بينهما هو أن
- المعامل المستخدم

يجب عرضه بشكل منفصل.

48. نتيجة الموجة والسلسلة الزمنية

لكل موجة أساسية معرف نتيجتها الخاص. يتم الاحتفاظ بنتيجة W1، نتيجة W2، نتيجة W3 بشكل منفصل. النتيجة الحالية: يمكن أن تكون آخر موجة أساسية أو مراجعة صالحة موجة إعادة اختبار. لا يجب أن يحدد متوسط الموجات السابقة النتيجة الحالية تلقائيًا.

49. النتيجة المتدحرجة

إذا رغب في ذلك، يمكن إنشاء نتيجة متحركة ثابتة زمنياً:

```
NOMOS_rolling= Σ_w δ_wNOMOS_w
```

هنا:

```
Σ_w δ_w= 1
```

هي. δw:

- قبل النتيجة،
- وفقًا لوزن الوقت

يجب تحديده. النتيجة المتدحرجة:

- النتيجة الحالية،
- لا تحل محل سجل الحوادث التاريخي

ولا يمكن لأي سجل أن يتم استبداله بصمت.

50. ثلاث درجات للادعاءات غير المحلولة

يمكن للطلبات غير المحلولة أن تزيد أو تقلل تقدير النقاط للذرات المادية بشكل مصطنع. لذلك، يُقترح ثلاث نتائج.

50.1. الحد الأدنى المحافظ

تعتبر الادعاءات غير المحلولة صفرًا. أبطأ

50.2. تقدير النقاط للمطالبات المحلولة فقط

يتم استخدام الادعاءات المحلولة فقط. Spoint

50.3. الحد الأعلى المتفائل

يتم قبول الادعاءات غير المحلولة بالدعم الكامل. Supper لذلك:

```
S_lower≤ S_point≤ S_upper
```

يجب أن يكون هذا هو الحال. هذا النطاق:

```
هو نطاق عدم اليقين المرجعي أو الإبستيمولوجي.
```

إنه ليس نفس شيء فترة الثقة للعينة.

51. عدم اليقين في العينة

نظرًا لأن تقديرات السكان تأتي من العينة:

- توزيع الاستجابة
- درجة المكون
- الدرجة المركبة

يجب أن يحمل فترة ثقة. الطريقة في العينات المعقدة:

- طبقات,
- المجموعات،
- أوزانهم،
- المستخدمون المزدوجون،
- أمواج

يجب حمايته. الطرق المقترحة:

- تقدير التباين القائم على التصميم
- أوزان التكرار
- الإقلاع الطبقي
- الإقلاع العنقودي
- نموذج الحدث النادر المناسب
- تقدير المسح بمساعدة النموذج

يجب تحديد الطريقة المستخدمة في السجل العام.

52. فترة الثقة للنتيجة المركبة

لا يمكن الحصول على فترة الثقة للنتيجة المركبة بمجرد جمع فترات الثقة للمكونات. المكونات:

- نفس المستخدمين،
- نفس الادعاءات،
- نفس اللغات،
- نفس الموجات

يعتمدون على هذا السبب. النهج المفضل: إنشاء عينات جديدة وفقًا لتصميم العينة. في كل عينة جديدة، أعد حساب جميع مقاييس الاستجابة والمكونات. أعد إنتاج الدرجة المركبة. قم بتوليد فترة الثقة من التوزيع. يجب أن تعيد هذه العملية تشغيل كامل هيكل الدرجات.

53. حجم العينة الفعّال

الملاحظة الخام: n والعينة الفعالة:

```
n_eff= (Σ_i w_i)^2/ Σ_i w_i^2 [K08]
```

يجب نشرها معًا. قد لا تعني الملاحظات الخام 1,000 وحدات المعلومات المساوية لـ .1,000 يجب أن تحمل درجات المكونات والفئات الفرعية أحجام العينات الفعالة الخاصة بها.

54. خطأ حرج تم ملاحظته صفر

عدم وجود أحداث حرجة مؤكدة في العينة:

تمت ملاحظة حرجة مؤكدة = 0

ينبغي أن يُكتب. يجب عدم استخدام التعبير التالي:

> الخطر الحرج هو صفر.

بافتراض الملاحظات المستقلة والمتساوية، حد أعلى تقريبي سريع:

```
p_upper,95≈ 3/n
```
[K09]

يمكن تقديره باستخدام النهج المعروف. في التصميم المعقد GEO-1000:

- بدلاً من n الخام، ينبغي استخدام طريقة مناسبة للتصميم،
- إذا لزم الأمر، عينة فعالة،
- حد أحادي الجانب دقيق أو قائم على الإعادة العشوائية (Bootstrap)

ينبغي استخدامه. مثال:

```
neff=1,000
```

والحد الأعلى التقريبي في ظل وجود صفر أحداث:

```
3/1,000 = 0.003 = 0.3%
```

قد يكون. البيان العام الصحيح:

> في ملاحظات *1,000* الفعالة، لم يتم ملاحظة أي أحداث حرجة مؤكدة؛ في الطريقة التقريبية المستخدمة، الحد الأعلى الأحادي الجانب بالنسبة المئوية لـ *95* يقارب *0.3* بالمئة.

55. عدم اليقين بالنسبة لحدث حرجة واحد

عند ملاحظة حدث حرجة واحد: قد يكون التقدير النقطي منخفضًا، وقد يكون مجال الثقة واسعًا. النتيجة الصحيحة:

- وَقَع الحدث،
- الانتشار المُقدّر،
- عدم اليقين الواسع،
- حالة الحدوث

يجب أن تُعرض معًا. الجملتان التاليتان قد تكونان خاطئتين أيضًا:

- «حدث واحد فقط غير مهم.»
- «الحدث يثبت أن النظام بأكمله قد فشل .100%»

56. حساسية الدرجة

يمكن أيضًا حساب البدائل التالية:

- الدرجة الخام غير الموزونة
- الدرجة المرجحة للسكان الرئيسيين
- درجة الوزن غير المحصّلة
- درجة الوزن المقتص
- الدرجة حسب الحلول فقط
- النتيجة المحافظة
- درجة المنتج المتساوي

- درجة مُوزونة حسب التعرض
- درجة نطاق الوصول الطبيعي
- درجة قاعدة المقارنة المشتركة
- درجة نظيفة
- الدرجة الطبيعية

إذا كان هناك فرق جوهري:

تحذير حساسية النتيجة

يجب تقديمه.

57. دقة غير صحيحة

قد يكون التمثيل التالي دقيقًا بشكل مفرط بالنسبة لمعظم نتائج GEO-1000: /963,742 1000 التقدير النقاطي:

- عدد صحيح،
- عشري إذا لزم الأمر

يجب نشره. يجب أن تشير فترة الثقة إلى حد الدقة العددية.

58. مكون مفقود

إذا لم يكن من الممكن حساب أحد المكونات العشرة:

- لا يمكن توزيع وزنه على المكونات الأخرى.
- لا يمكن احتساب الحقل المفقود كنقاط 100،

لا يمكن نشر النتيجة الكاملة NOMOS. الحالة الصحيحة:

ملف **NOMOS** جزئي

المكوّن غير قابل للتقدير

النتيجة التقديرية

قد تنطبق. مثال: إذا لم يكن هناك لوحة طبيعية لـ USI متاحة، يمكن عرض ملف جزئي يحتوي على ثمانية أو تسعة مكونات فقط.

```
لا يمكن افتراض USI=100.
```

59. حالة نشر النتيجة

NSQ-0— لم يتم الحساب

البيانات المطلوبة غير متوفرة.

NSQ-1— تقديرية

موجة واحدة، تحكيم محدود، أو وجود مكونات مفقودة.

—NSQ-2 ملف جزئي

تم حساب بعض المكونات بشكل موثوق. لا يمكن إجراء مقارنة كاملة لنقطة 1,000.

—NSQ-3 نتيجة موجة صالحة

تم استيفاء جميع المكونات الأساسية والمتطلبات المسبقة للجودة.

—NSQ-4 نتيجة مكررة

تم تكرار النتيجة في موجة مستقلة واحدة على الأقل أو موجة تأكيدية.

—NSQ-5 نتيجة تم تدقيقها خارجيًا

تم إعادة إنتاج الحساب وسلسلة الأدلة بواسطة فريق مستقل. مؤهلية +950 على الأقل:

NSQ-4

يجب أن يستهدف ذلك المستوى.

60. إصدارات الدرجات

يجب أن يشمل كل إصدار للدرجة ما يلي:

- نسخة كود الحساب
- صيغ المكونات
- تعيينات قيم الأبعاد
- أوزان عائلة صياغات المُدخل
- معاملات المنتج
- أوزان العينات
- نسخة حزمة الحقيقة
- نسخة المحكم
- معرّف الموجة
- هاش النزاهة

إذا تغيرت الصيغة: لا يمكن إعادة كتابة النتيجة القديمة، نسخة النتيجة الجديدة يُنشأ.

61. صيغة الدرجة الجديدة لنفس البيانات

قد تتطور منهجية التسجيل. نفس البيانات الخام:

- طريقة التسجيل 0.9
- طريقة التسجيل 1.0

يمكن إعادة حسابها مع. السجل الصحيح:

البيانات	طريقة التسجيل	النتيجة
مجموعة الملاحظات الثابتة W1	0.9	944
مجموعة الملاحظات الثابتة W1	1.0	951

تعتبر النتيجتان إصدارات منفصلة. لا يمكن للطريقة الجديدة حذف الطريقة القديمة من الماضي.

62. الأبواب ضد التلاعب في التسجيل

يجب أن تحتوي بنية التسجيل NOMOS على حماية صريحة ضد الطرق التالية:

- اختيار منتجات الذكاء الاصطناعي ذات التسجيل العالي فقط
- استبعاد اللغة أو الدولة ذات التسجيل المنخفض من النطاق
- تحويل اجتياز مشروط إلى اجتياز كامل
- اعتبار غير المحلولة إيجابية
- إزالة غير قابل للتقييم من المقام بصمت
- تغيير وزن عائلة المُدخل بعد النتيجة
- خفض معامل المنتج الكبير حسب النتيجة
- إخفاء المنتج ذو الحدث الحرج في متوسط النظام البيئي
- نشر النتيجة النظيفة فقط
- اختيار أفضل موجة كدرجة حالية
- إخفاء الفرق بين الدرجات غير الموزونة والموزونة
- توزيع وزن المكونات المفقودة على المكونات القوية
- عدم توفير فاصل الثقة
- تجاوز عتبة 950 بناءً على التقدير النقطي وإخفاء حالة عدم اليقين

63. بطاقة النتيجة العامة NOMOS

يجب أن تحتوي بطاقة النتيجة العامة للمرشح على الترتيب التالي:

أ. الهوية

الكيان المدقق منتج AI نطاق الوصول الطبيعي/ قاعدة المقارنة المشتركة تم التحكم فيه / مجموعة الإرشادات الطبيعية إصدار Wave Truth Pack طريقة التقييم

ب. حالة البوابة

الامتثال الكامل للمرشح الشرطي إخفاق جسيم تعليق حرج غير قابل للتقييم

ج. توزيع الاستجابة **GEO-1000**

RP-1–RP-8 معادلات المستخدمين المرور الصارم المرور المقبول حرج رئيسي غير محلول لا يوجد استجابة قابلة للاستخدام غير قابل للتقييم

د. ملف المكون العشرة

ENT

FAC

SBI

TLA

ECI

EPI

TCR

LGF

USI

STR

هـ. النتيجة المركبة

NOMOS/ 1,000
فاصل الثقة في العينة المرجع الحدود الدنيا/القصوى العينة الفعالة

و. المخاطر والأحداث
الأحداث الحرجة المؤكدة المرصودة؛ الحد الأعلى الحرج؛ النتائج الجسيمة؛ الحوادث المفتوحة؛ وحالة التصحيح.

ج. النطاق والجودة

SQ

PSI

NCL

TPR

AQ

NSQ
تغطية الدولة/اللغة مستوى المقارنة

.64 توزيع الاستجابات الاصطناعية

عرض منهجية الاصطناع / النتائج التالية ليست منتجات فعلية للذكاء الاصطناعي أو أداء الشركة. بالنسبة لـ Synthetic Orion AI:

الحالة	ما يعادل المستخدم
RP-1 اجتياز كامل	730
RP-2 اجتياز مع ملاحظة	170
RP-3 اجتياز مشروط	70
RP-4 إخفاق جسيم	20
RP-5 إخفاق حرج	2
RP-6 لم يتم حله	5
RP-7 لا توجد استجابة قابلة للاستخدام	3
RP-8 غير قابل للتقييم	0
الإجمالي	1,000

اجتياز صارم:

```
730+170=900
```

اجتياز مقبول:

```
900+70=970
```

حرج: 2/1,000 كبير: 20/1,000 نسبة الاجتياز المقبول لهذا المنتج هي 97 في المئة. ومع ذلك، هناك نتائج حرجة وكبيرة واضحة. الحالة الصحيحة:

تعليق حرج

الحالة غير الصحيحة:

97 ناجح بنسبة مئوية

65. الملف التوليفي للمركب الاصطناعي

المكون	الدرجة
ENT	96
FAC	93
SBI	89
TLA	95
ECI	88
EPI	91
TCR	96
LGF	82
USI	87
STR	90
NOMOS	907

العرض العام الصحيح:

تعليق حرج— 907/1,000

هذا النظام:

- في المتوسط، مناطق قوية,
- ضعف اللغة/عدالة البلد،
- قضايا الأدلة المادية،
- حدث حرج مفتوح

يحمل.

66. نتيجة التصحيح التركيبية

في الموجة الجديدة:

الحالة	ما يعادل المستخدم
RP-1	830
RP-2	128
RP-3	30
RP-4	0
RP-5	0
RP-6	5
RP-7	7
RP-8	0
الإجمالي	1,000

اجتياز صارم: مكونات :958

المكون	الدرجة
ENT	99
FAC	97
SBI	96
TLA	95
ECI	96
EPI	98
TCR	97
LGF	94
USI	93
STR	96
NOMOS	961

السجل الصحيح: الحالي: NOMOS +950 مرشح الأداء — 961/1,000/ التاريخي: الحادث الحرج المؤكد السابق تم تداركه للمرشح / الحرج الملاحظ في إعادة الاختبار الحالية: 0/ الحد الأعلى لمعدل الحرج: يظهر أيضًا الحادث الحرج القديم لم يُحذف.

67. حالة معامل منتج اصطناعي للذكاء الاصطناعي

أربعة منتجات ذكاء اصطناعية:

المنتج	NOMOS	معامل متساوٍ	معامل التعرض
A1	960	0.25	0.55
A2	940	0.25	0.25
A3	900	0.25	0.15
A4	980	0.25	0.05

درجة المنتج المتساوية:

```
0.25(960+940+900+980)=945
```

الدرجة الموزونة حسب التعرض:

```
0.55(960)+0.25(940)+0.15(900)+0.05(980)=947
```

النتيجتان متقاربتان. ومع ذلك، يمكن أن تكونا مختلفتين جدًا في توزيع آخر. التفسير الصحيح:

- درجة النظام البيئي المتساوي: 945
- درجة النظام البيئي الموزونة حسب التعرض: 947
- سجلات البوابة والحوادث على مستوى المنتج منفصلة

تعليق غير صحيح: "درجة النظام البيئي لـ NOMOS هي بالتأكيد 947."

68. حالة معامل حرجة تركيبية

لتكن درجة المنتج A4 هي 980. ومع ذلك، لنفترض وجود حدث حرِج مؤكد في A4. معامل التعرض هو فقط 0.05. قد يظل المركب البيئي مرتفعًا. البطاقة العامة الصحيحة: المركب المرجح حسب التعرض: 947/ أي منتج حرِج مؤكد: نعم / حصة التعرض الحرِج: النسبة المحسوبة / تغطية A4: تعليق حرج لا يمكن للمعامل الصغير لـ A4 إزالة الحدث الحرِج.

69. قرار حالة نشر الدرجات

ترتيب قرار المرشح:

```
PublicationStatus= غير قابل للتقييم(القياس الأساسي مفقود); تعليق حرج(مفتوح مؤكد حرج);
إخفاق جسيم(مفتوح مؤكد رئيسي); PROVISIONAL(المكون أو التكرار أو الجودة مفقود);
CONDITIONAL(متوسط أو عتبة غير محلولة); 950+ CANDIDATE(جميع متطلبات المرشح);
وإلا صالحة
```

يتم عرض نطاق النتيجة بعد قرار الحالة هذا.

70. الأحكام الإلزامية المعيارية

CH16-N01

لا يمكن نشر نتيجة NOMOS كنقطة مركبة منفردة.

CH16-N02

يجب عرض حالة البوابة قبل النتيجة المركبة.

CH16-N03

توزيع استجابة GEO-1000 هو جزء إلزامي من النتيجة العامة الرئيسية.

CH16-N04

يجب أن يكون مجموع ما يعادل المستخدمين RP-1 وRP-8 هو 1,000.

CH16-N05

يجب عرض النجاح الصارم، النجاح المقبول، الرئيسي، الحرج، غير المحلول، لا توجد استجابة قابلة للاستخدام، و غير قابل للتقييم بشكل منفصل.

CH16-N06

لا يمكن إضافة الملاحظات غير المحلولة و غير قابل للتقييم بصمت إلى فئات النجاح أو الفشل.

CH16-N07

لا يمكن تفسير ملاحظات غير قابل للتقييم على أنها فشل في أداء المنتج.

CH16-N08

لا يمكن إزالة ملاحظات غير قابل للتقييم بصمت من المقام؛ يجب عرضها في التغطية القابلة للتقييم.

CH16-N09

يجب أن تكون أوزان الملاحظات مستقلة عن نتيجة استجابة الذكاء الاصطناعي.

CH16-N10

يجب الإبلاغ عن العدد الخام للمشاركين وما يعادله للمستخدمين الموزون حسب السكان بشكل منفصل.

CH16-N11

يجب الاحتفاظ بحالة اجتياز الاستجابة بشكل منفصل عن نسبة دعم الادعاء الذرية.

CH16-N12

لا يمكن تعويض البوابات الحرجة والجسيمة بأي درجة مركبة.

CH16-N13

يجب تحديد أوزان المركزية الذرية قبل الاطلاع على النتائج.

CH16-N14

لا يمكن لتكرار نفس الادعاء الجذرية زيادة أو تقليل النتيجة بشكل مصطنع.

CH16-N15

لا يمكن لأوزان مركزية الادعاء استبدال البوابات الحرجة والجسيمة.

CH16-N16

يجب توثيق قيم بعد الأمانة الذرية وإتاحتها علنًا بالإصدار.

CH16-N17

إذا تغيرت خريطة الأمانة الذرية، يجب إنشاء نسخة جديدة لطريقة التقييم.

CH16-N18

لا يمكن إضافة ذرات فجوة مرجعية بصمت إلى درجة النقاط كصفر أو واحد.

CH16-N19

يجب نشر اكتمال عنصر الاستجابة المطلوب كمقياس منفصل.

CH16-N20

لا يمكن تعويض فشل العنصر الأساسي المطلوب بمتوسط REC مرتفع.

CH16-N21

يجب أن يكون الملف الشخصي NOMOS المكون من عشرة عناصر مرئيًا قبل المؤشر المركب.

CH16-N22

يجب أن تحمل ENT و FAC و SBI و TLA و ECI و EPI و TCR و LGF و USI و STR درجات منفصلة.

CH16-N23

لا يمكن لمكون واحد أن يتولى بهدوء الوزن المفقود لمكون آخر.

CH16-N24

لا يمكن افتراض أن المكون الذي لا يمكن حسابه هو 100.

CH16-N25

يجب نشر المؤشر المركب الكامل NOMOS فقط إذا تم حساب جميع العناصر العشرة بشكل موثوق.

CH16-N26

إذا تم استخدام الوزن المتساوي للمكونات كطريقة مرجعية، يجب ذكر ذلك صراحة.

CH16-N27

يجب أن يكون للأوزان البديلة الخاصة بالمجال اسم نسخة منفصل.

CH16-N28

لا يمكن استخدام درجة خاصة بالمجال مكان المؤشر المرجعي بدون تفسير.

CH16-N29

لا يمكن لمؤشر LGF الاعتماد فقط على متوسط السكان؛ يجب أن يشمل قاعدة الأداء والفروق.

CH16-N30

لا يمكن تحويل ملاحظة واحدة من مراقب الدولة إلى درجة عدالة الدولة.

CH16-N31

يجب أن تشارك فقط الخلايا اللغوية والجغرافية التي تمتلك حجم عينة كافٍ في تقدير العدالة.

CH16-N32

لا يمكن تجاهل اللغات والدول التي لم يتم اختبارها من أجل تحقيق نتيجة عادلة عالية.

CH16-N33

يجب على USI تقييم نتائج اللوحات المسيطر عليها والطبيعية معًا بالإضافة إلى فرق واجهة المادة.

CH16-N34

في دراسة تحتوي على لوحة مسيطرة فقط أو لوحة طبيعية فقط، لا يمكن اعتبار USI محسوبًا بالكامل.

CH16-N35

يجب الاحتفاظ بنتائج اللوحات المسيطر عليها والطبيعية منفصلة بشكل افتراضي.

CH16-N36

إذا تم دمج نتائج اللوحات المسيطر عليها والطبيعية، يجب تثبيت المعاملات قبل النتائج.

CH16-N37

لا يمكن تقديم STR كدرجة استقرار كاملة في موجة واحدة.

CH16-N38

يجب نمذجة التغيرات بين الموجات للمنتج أو Truth Pack صراحة في حساب الاستقرار.

CH16-N39

المركب NOMOS لا يمكنه تغيير حالة البوابة الحرجة أو الرئيسية.

CH16-N40

لا يمكن لنقطة واحدة من 950 وحدها أن تثبت ترشيح NOMOS +950.

CH16-N41

يجب أن تلبي شروط المرشح NOMOS +950 جميعها معًا متطلبات النقطة، النجاح الصارم، البوابة، مستوى المكون، غير المحلول، الجودة، وشروط التكرار.

CH16-N42

لا يمكن تقديم حدود +950 كمعيار علمي نهائي قبل التجربة التجريبية والمراجعة المستقلة.

CH16-N43

يجب أن تحمل هوية النقطة منتج الذكاء الاصطناعي، الإطار، اللوحة، مجموعة صياغات المُدخل، الموجة، حزمة الحقيقة، وإصدار النقطة.

CH16-N44

لا يمكن استخدام الجملة "نقطة NOMOS للكيان هي X" بدون هوية النطاق.

CH16-N45

لا يمكن عرض نتيجة GEO-1000 Core كدرجة NOMOS Full.

CH16-N46

يجب الإبلاغ عن درجات نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة بشكل منفصل.

CH16-N47

لا يمكن تصنيف درجات المنتج لفئات نطاق الوصول الطبيعي المختلفة مباشرة دون توضيح.

CH16-N48

لا يمكن عرض نتيجة قاعدة المقارنة المشتركة كدرجة الوصول العالمي الكاملة للمنتج.

CH16-N49

يجب تحديد معاملات المنتج للذكاء الاصطناعي بشكل مستقل عن النتائج.

CH16-N50

لا يمكن أن يعتمد معامل المنتج للذكاء الاصطناعي على درجة المنتج أو سمعة العلامة التجارية أو علاقة الرعاية التجارية.

CH16-N51

يجب تسمية درجات النظام البيئي المتساوية للمنتج والمحسوبة على أساس التعرض بشكل منفصل.

CH16-N52

إذا لم يكن من الممكن التحقق من بيانات المستخدم النشط أو التعرض، لا يمكن توليد المعاملات التعسفية.

CH16-N53

يجب أن تكون بيانات التعرض شفافة على مستوى الحساب أو المستخدم أو الاستعلام أو الجلسة.

CH16-N54

لا يمكن خفض معامل منتج منخفض النقاط بعد رؤية النتيجة.

CH16-N55

لا يمكن لمركب النظام البيئي حذف بوابة حرجة أو جسيمة على مستوى المنتج.

CH16-N56

يجب عرض نسبة Critical وnyCriticalA المرجحة بالتعرض بشكل منفصل.

CH16-N57

يجب تحديد معاملات عائلة المُدخلات قبل النتائج.

CH16-N58

لا يمكن تقليل وزن عائلة المُدخلات منخفضة النقاط لاحقًا.

CH16-N59

يجب أن تظل النتيجة المنفصلة لكل عائلة من صياغات المُدخل مرئية جنبًا إلى جنب مع درجة صياغة المُدخل المركبة.

CH16-N60

يجب الاحتفاظ بدرجة الموجة الحالية منفصلة عن الدرجات المتداولة أو التاريخية.

CH16-N61

لا يمكن اختيار الموجة السابقة ذات الدرجة الأعلى كدرجة حالية.

CH16-N62

يمكن إنتاج النتائج المحافظة، ونتائج الحلول فقط، والمتفائلة للكتلة غير المحلولة من الادعاءات.

CH16-N63

لا يمكن تقديم عدم اليقين المرجعي وفاصل الثقة للعينة على أنهما الشيء نفسه.

CH16-N64

يجب أن يحافظ فاصل الثقة للعينة المعقدة على التوزيع الطبقي، والتكتل، والوزن، وبنية المستخدم المطابقة.

CH16-N65

يجب حساب فترة الثقة للدرجة المركبة عن طريق إعادة تشغيل عملية التقييم بأكملها.

CH16-N66

يجب عرض أحجام العينات الخام والفعالة معًا.

CH16-N67

لا يمكن التعبير عن حدث حرج يُلاحظ على أنه صفر كمخاطر حرجة صفرية.

CH16-N68

في حالة عدم وجود أحداث، يجب نشر حد علوي من جانب واحد أو عدم اليقين المعادل للأحداث النادرة.

CH16-N69

لا يمكن استخدام نهج n3/ تلقائيًا كطريقة نهائية في التصميم المعقد؛ يمكن أن يكون فقط تحكمًا تقريبيًا.

CH16-N70

في حدث حرج، يجب إظهار وجود الحدث وعدم اليقين بشأن تكراره معًا.

CH16-N71

تحليلات حساسية الدرجات يجب أن تظهر النتائج غير الموزونة، والموزونة، والمقصوصة، وغير المحلولة، والنتائج البديلة للمعامل بالمدى ذي الصلة.

CH16-N72

حساسية الدرجة الجوهرية يجب أن تخلق تنبيهًا على البطاقة العامة.

CH16-N73

لا يمكن نشر الكسور العشرية المزيفة للدرجة على الإطلاق.

CH16-N74

لا يمكن توزيع وزن المكون المفقود على المكونات الأخرى.

CH16-N75

يجب أن يكون حالة جودة نشر درجة Puan مرئية في السجل العام.

CH16-N76

التقدم للترشح +950 يجب أن يهدف على الأقل إلى مستوى الدرجة المكررة.

CH16-N77

عندما تتغير صيغة الدرجة، لا يمكن إعادة كتابة الدرجات السابقة بصمت.

CH16-N78

إذا تم تطبيق نسخ مختلفة من طريقة الدرجات على نفس مجموعة البيانات، يجب الحفاظ على جميع النتائج في نسخ منفصلة.

CH16-N79

يجب نشر كود حساب الدرجات والمعاملات وسجلات السلامة بتنسيق قابل للقراءة آليًا.

CH16-N80

يجب أن يكون لكل درجة، ومعامل، وفاصل ثقة، وقرار نشر إنسان أو مؤسسة مسؤولة وملزمة.

71. أشكال الفشل

CH16-F01 — تقرير رقم واحد

يتم الحفاظ على التوزيع، البوابة، وملف المكونات.

CH16-F02 — إخفاء البوابة تحت الدرجة

تعليق حرج تصبح حاشية صغيرة.

CH16-F03 — عد الشرطية كـ اجتياز كامل

المشكلات المعتدلة تصبح غير مرئية.

CH16-F04 — عد غير المحلولة كنجاح

النتائج غير المعروفة ترفع النتيجة.

CH16-F05— عد غير المحلولة كفشل
نقص المرجع يقلل المنتج.
CH16-F06— إزالة غير قابل للتقييم من المقام
عيوب القياس تصبح غير مرئية.
CH16-F07— عد غير قابل للتقييم كعيب في المنتج
عيب الفحص يُسند إلى الذكاء الاصطناعي.
CH16-F08— عد 1,000 الأشخاص بالتساوي في الوزن
يتم تجاهل الإفراط في العينات والفروق السكانية.
CH16-F09— الوزن وفقًا للنتيجة
الإجابات الصحيحة تحصل على وزن أكبر.
CH16-F10— احسب معدل دعم الادعاءات كإجابة تمر
البوابة الحرجة تضيع في المتوسط.
CH16-F11— احتساب الأساسي والفرعي على قدم المساواة
الادعاء العرضي يحصل على نفس وزن الكيان الرئيسي.
CH16-F12— تضخيم النتيجة بالتكرار
يتم كتابة نفس الادعاء الصحيح عشر مرات لزيادة النتيجة.
CH16-F13— تقليص النتيجة بالتكرار
الجذر الخطأ نفسه يؤدي إلى عشرة عقوبات منفصلة.
CH16-F14— تخزين قيم الأبعاد
لا يمكن إعادة إنتاج النتيجة.
CH16-F15— عد الادعاء القادم من الأسفل كحقيقة لـ F-4
الانتقال الدقيق للتقرير الذاتي يتحول إلى دقة التقرير الذاتي.
CH16-F16— إذابة العناصر المطلوبة في المتوسط الذري
حتى إذا كانت المهمة الرئيسية مفقودة، يتم منح الإجابة درجة عالية.
CH16-F17— انتقاء المكونات بعناية
يتم نشر المكونات الثلاثة القوية فقط.
CH16-F18— عد المكون المفقود كـ 100
تم زيادة الدرجة الكاملة بشكل مصطنع.
CH16-F19— توزيع الوزن المفقود
تتم إضافة درجة العدالة غير المحسوبة إلى الدرجة الواقعية.
CH16-F20— عد قانون الطبيعة على أنه وزن متساوٍ
لا يتم إجراء الحساسية أو المراجعة الخارجية.
CH16-F21— عد درجة المجال كدرجة معيارية
يتم تقديم الوزن الخاص مثل درجة NOMOS الأساسية.
CH16-F22— اعتبار متوسط اللغة كعدالة لغوية

يتم إخفاء الحد الأدنى للغة.

CH16-F23— احتساب تساوي اللغة كجودة

الدرجات المنخفضة في جميع اللغات تنتج عدالة عالية.

CH16-F24— احتساب اللغات غير المختبرة كجزء من التغطية

عامل التغطية مضخم.

CH16-F25— درجة الدولة من المراقب

يدخل مستخدم واحد في حساب العدالة.

CH16-F26— نشر الدرجة النظيفة فقط

يتم الحفاظ على تجربة المستخدم الطبيعية.

CH16-F27— احتساب الدرجة الطبيعية فقط كإعداد افتراضي للمنتج

تصبح التخصيص السلوكي الأساسي.

CH16-F28— هادئ بشكل متوسط ونقي وطبيعي

تختلط أهداف التنبؤ المختلفة.

CH16-F29— 100 الاستقرار في موجة واحدة

يُعتبر الاستقرار كاملاً دون تكرار.

CH16-F30— عد الموجة الجيدة القديمة كموجة حالية

يتم إخفاء الضعف الحالي.

CH16-F31— عد 950 تلقائيًا باعتباره امتثالًا

يتم تجاهل البوابة وطابق المكونات.

CH16-F32— تعليق حرج— عد 970 كـ +950

العد يسبق البوابة.

CH16-F33— إخفاء هوية الدرجة

لا يُعرف أي منتج ولوحة يتم قياسهما.

CH16-F34— احتساب الدرجة الأساسية كدرجة كاملة NOMOS

يمثل مُدخل واحد المعيار بأكمله.

CH16-F35— ترتيب نتائج نطاق الوصول الطبيعي

يُعتبر السكان المختلفون متشابهين.

CH16-F36— احتساب الوصول الكامل إلى قاعدة المقارنة المشتركة

يتم تحويل مجموعة سكانية محدودة إلى نتيجة عالمية.

CH16-F37— إعطاء معامل النموذج وفقًا للدرجة

تحصل الدرجة العالية على تأثير أكبر.

CH16-F38— جعل سمعة العلامة التجارية عاملاً

يتم وزن شعبية المنتج دون التنبؤ بها.

CH16-F39— احتساب أرقام الحسابات كمستخدمين نشطين

معامل التعرض يتضخم.

CH16-F40— إزالة المنتجات منخفضة النتيجة
يتكون النظام البيئي فقط من المنتجات الناجحة.
CH16-F41— احتساب درجة التعرض المتساوية للمنتج
يتم عرض المنتج المتوسط مثل الشخص المتوسط.
CH16-F42— احتساب درجة التعرض كجودة للمنتج
المنتج الكبير يصبح تلقائيًا مقياس جودة أكثر أهمية.
CH16-F43— حذف العنصر الحرج في المتوسط البيئي
حدث كبير للمنتج ذو المعامل الصغير يختفي.
CH16-F44— تغيير معامل صياغة المُدخل بعد النتيجة
العائلة ذات صياغة المُدخل الصعبة يتم تقليصها.
CH16-F45— اعتبر العائلة منخفضة النتيجة اختيارية
الحد أو نتيجة الدليل مستبعدة من النطاق.
CH16-F46— الحفاظ على الحال مع النتيجة المتغيرة
الأمواج العالية القديمة تغطي الانخفاض الحالي.
CH16-F47— حذف العناصر غير المحلولة من المقام
التقدير النقطي يرتفع بشكل مصطنع.
CH16-F48— النظر في عدم اليقين المرجعي في ثقة المعاينة
يختلط نوعان مختلفان من عدم اليقين.
CH16-F49— ثقة العينة العشوائية البسيطة
يتم تجاهل الطبقات، الوزن، والعناقيد.
CH16-F50— جمع ثقات المكونات
يتم تجاهل الاعتماد.
CH16-F51— ادعاء الحساسية بعدد N الخام
يتم إخفاء حجم العينة الفعلي.
CH16-F52— احتساب الأحداث صفر على أنها خطر صفر
يتم إخفاء عدم اليقين في الأخطاء الحرجة النادرة.
CH16-F53— تطبيق /N3 على كل حالة
تم تجاهل هيكل العينة والوزن المعقد.
CH16-F54— عد حالة حرجة واحدة كـ 100 لكل نظام نسبة المخاطر
التعميم الشائع مبالغ فيه.
CH16-F55— إخفاء حالة حرجة واحدة لأن المعدل صغير
اختفاء وجود الحدث.
CH16-F56— إخفاء تحليل الحساسية
لا يُرى أن النتيجة تتغير بالطرق البديلة المعقولة.
CH16-F57— حساسية عشرية مزيفة

تم نشر نتيجة مثل 963,742.
CH16-F58— حساب الملف الجزئي كدرجة كاملة
يتم إجراء المقارنة مع مكون مفقود.
CH16-F59— نشر النتيجة المؤقتة باعتبارها نهائية
موجة واحدة وقاضٍ واحد يصبحان المعيار الكامل.
CH16-F60— إعادة كتابة التاريخ في تغيير الصيغة
النتيجة القديمة تختفي.
CH16-F61— إبقاء كود الحساب مغلقًا
يتم منع إعادة الإنتاج المستقلة.
CH16-F62— اعتبار ارتفاع النتيجة كجودة تدقيقية
يتم قبول قياس خاطئ لأنه أعطى نتيجة عالية.
CH16-F63— اعتبار انخفاض النتيجة كعيب في الطريقة
يتم إعادة حساب النتيجة المنخفضة بسبب الإزعاج التجاري.
CH16-F64— احتكار ناتج شركة واحدة
يتم منع المؤسسات المستقلة من تطبيق نفس الصيغة.

72. إجراء التدقيق

الخطوة 1— إنشاء هوية النتيجة

تم تثبيت طريقة تسجيل الدرجات لمجموعة كائنات منتج الذكاء الاصطناعي من لوحة الإطار والمُدخل ومجموعة الحقيقة المعبأة.

الخطوة 2— تثبيت مجموعة الملاحظة التحليلية الرئيسية

يتم فصل النسخ الاحتياطية والمسحوبات والسجلات غير القابلة للتقييم والسجلات الرئيسية للملاحظة.

الخطوة 3— التحقق من أوزان السكان

يتم فحص الأوزان الخاصة بالتصميم وعدم الاستجابة والمعايرة.

الخطوة 4— حساب توزيع RP

يتم إنتاج ما يعادل المستخدم مع RP-1 إلى RP-8.

الخطوة 5— حساب الصلاحية الصارمة والمقبولة

يتم اشتقاق SP1 000 و AP1 000.

الخطوة 6— حساب كثافات الحرجة، الكبرى، غير المحلولة، وغير القابلة للتقييم

يتم الاحتفاظ بالبوابة والانتشار بشكل منفصل.

الخطوة 7— تطبيق مركزية الادعاءات وتطبيع التجمع الجذري

يتم التحقق من الادعاءات مرة أخرى.

الخطوة —8 حساب مقاييس الحجم الذري

يتم إنتاج نتائج الكيان والحقيقة والنطاق والوقت والنسبة والطريقة والمرجع.

الخطوة —9 حساب اكتمال العنصر المطلوب

يتم تسجيله مع بوابات عائلة المطالب.

الخطوة —10 حساب عشرة مكونات

يتم تطبيق جميع صيغ المكونات من ENT إلى STR.

الخطوة —11 التحقق من المكونات المفقودة

يتم تحديد ما إذا كان يمكن حساب الدرجة الكاملة.

الخطوة —12 حساب مؤشر NOMOS المركب

ينطبق فقط على الملفات الشخصية الكاملة والصحيحة.

الخطوة —13 تعيين حالة البوابة

يحدد ما إذا كانت الحالة حرجة، كبيرة، مشروطة، أو مرشح كامل.

الخطوة —14 فصل نتائج نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة

تُحفظ مجموعتا الدرجات بشكل منفصل.

الخطوة —15 تثبيت معاملات منتج AI

يتم حفظ معاملات التساوي والتعرض مع مصدر البيانات.

الخطوة —16 تطبيق معاملات عائلة prompt

تُستخدم الأوزان بشكل مستقل عن النتيجة.

الخطوة —17 فصل نتائج اللوحة والموجة

تُفصل الدرجات النظيفة والطبيعية والحالية والمتدرجة.

الخطوة —18 حساب مقاييس العدالة والحد الأدنى

يتم إنتاج نتائج اللغة والبلد والتغطية.

الخطوة —19 حساب نتائج النقاط السفلية والعلوية غير المحلولة

يتم توضيح عدم اليقين المرجعي.

الخطوة —20 حساب فترات الثقة للعينة

يتم تطبيق طريقة إعادة أخذ العينات المناسبة لتصميم العينة.

الخطوة —21 حساب حجم العينة الفعال

معروض على الأساس الكامل وأساس المكونات.

الخطوة —22 حساب عتبة الحدث النادر الحرج

يتم استخدام طريقة وحيدة الجانب للأحداث الصفرية أو القليلة.

الخطوة —23 توليد تحليلات الحساسية

يتم مقارنة الوزن والمعامل والخيارات غير المحلولة.

الخطوة —24 تعيين حالة جودة النشر

يتم اتخاذ قرار بين NSQ-0 و NSQ-5.

الخطوة —25 تطبيق NOMOS +950 أبواب المرشحين

يتم التحقق من جميع الشروط باستثناء النتيجة.

الخطوة —26 إنشاء بطاقة النتائج العامة

يتم نشر البوابة والتوزيع والملف الشخصي والفهرس وعدم اليقين والنطاق بترتيب معين على التوالي.

الخطوة —27 تثبيت رمز الحساب والبيان

تم ربط نسخة الشيفرة، المدخلات، التجزئة، والنتائج.

الخطوة —28 تحضير حزمة الاستنساخ المستقلة

تتوفر ملفات الحوسبة الخالية من البيانات الشخصية للمراجعة المستقلة.

73. الأدلة المطلوبة

تسجيل كيان معرف الهدف نظام الذكاء الاصطناعي نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة حالة اللوحة المجموعة المستجيبة والعائلات إصدار حقيبة الحقيقة معرف الموجة سجلات الادعاءات قرارات الاستجابة RP-1–RP-8 التوزيع مجموعة الملاحظة للتحليل الرئيسي أوزان التصميم أوزان عدم الاستجابة أوزان المعايرة تقليم الأوزان سجلات العينة الفعالة اجتياز صارم اجتياز مقبول معدل حرج معدل رئيسي معدل غير محلول معدل لا توجد استجابة قابلة للاستخدام معدل غير قابل للتقييم أوزان مركزية للادعاء سجلات مجموع الجذر قيم البعد الذري اتساق الذرة العنصر المطلوب اكتمال

ENT

FAC

SBI

TLA

ECI

EPI

TCR

LGF

USI

STR

مكونات المدخلات الفرعية؛ الحد الأدنى للغة والبلد؛ عامل التغطية؛ الدرجات المسيطر عليها والطبيعية؛ نتائج بين الموجات؛ مؤشر مركب NOMOS؛ درجات بديلة خاصة بالمجال؛ معاملات منتج الذكاء الاصطناعي ومصادر بياناتها؛ معاملات عائلة صياغات المُدخل؛ نتيجة المنتج المتساوي؛ النتيجة المرجحة بالتعرض؛ حالة AnyCritical؛ حصة التعرض الحرجة؛ النتائج غير المحلولة للنقطة السفلى–العليا؛ فترة ثقة المعاينة؛ الحد الأعلى الحرج؛ تحليل الحساسية؛ جودة نشر الدرجات؛ سجل بوابة المرشح +950؛ نسخة طريقة التقييم؛ كود الحساب؛ تجزئة سلامة الكود؛ بطاقة الدرجات العامة؛ سجل التغييرات؛ سجل الاستنساخ المستقل؛ والشخص أو المؤسسة المسؤولة.

74. قائمة تدقيق التدقيق

هل يغطي تعريف النتيجة جميع مجالات النطاق؟ هل تم تثبيت مجموعة الملاحظات الرئيسية للتحليل قبل النتيجة؟ هل تم فصل النسخ الاحتياطية والملاحظات الرئيسية؟ هل أوزان الملاحظات مستقلة عن النتيجة؟ هل توزيع RP-1–RP-8 هو مجموع 1,000؟ هل تم فصل اجتياز الصارم والمقبول؟ هل الأهمية الحرجة والكبيرة منفصلة؟ هل تم تعليم غير المحلولة كنجاح أم فشل؟ هل تم إزالة غير قابل للتقييم من المقام؟ هل التغطية القابلة للتقييم مفتوحة؟ هل استخدم معدل دعم الادعاءات بدل رد النجاح؟ هل أوزان المركزية مفتوحة؟ هل تم احتساب نفس الادعاء الأساسية مرة أخرى؟ هل تم نشر تخطيط البعد الذري؟ هل تم تفسير F-4 بشكل صحيح؟ هل تم حساب العناصر المطلوبة بشكل منفصل؟ هل تم تعويض البوابة الأساسية بزيادة REC؟

هل المكوّن بالكامل حاضر؟ هل تم احتساب المكوّن المفقود على أنه 100؟ هل تم توزيع وزن المكوّن المفقود؟ هل يقيس ENT العلاقات الصحيحة بين الكيانات؟ هل يستبعد FAC بهدوء الادعاءات غير المحلولة؟ هل يشمل SBI إغفال حدود المادة؟ هل يحافظ TLA على الفرق بين الدولة والوقت؟ هل يقيس ECI دعم الاستشهاد بدلاً من وجود الاستشهاد؟ هل يحمل EPI التمييز بين الإعلان الذاتي والواقع؟ هل يجعل TCR الاستجابة غير القابلة للاستخدام مرئية؟ هل يشمل LGF الأساس والفجوة إلى جانب المتوسط؟ هل أدخلت مراقبة الدولة نتيجة الدولة؟ هل يعتمد USI على كل من البيانات النظيفة والطبيعية؟ هل تم إنتاج STR من موجة واحدة؟ هل غطت النتيجة المركبة حالة البوابة؟ هل يعتمد +950 فقط على نتيجة النقاط؟

هل تم فحص المكونات في الطوابق؟ هل الحد غير المحلَّل للنواة مفتوح؟ هل تم فصل نطاق الوصول الطبيعي و قاعدة المقارنة المشتركة؟ هل تم تحديد معاملات منتج الذكاء الاصطناعي قبل رؤية الدرجات؟ هل تقيس المعاملات المستخدمين النشطين أم الحسابات أم الاستفسارات؟ هل نتائج المساواة والتعرض منفصلة؟ هل تمت إزالة المنتج

منخفض الدرجة؟ هل يخفي تقييم النظام البيئي المنتج الحرج؟ هل يوجد حقل nyCriticalA؟ هل أوزان عائلة صياغات المُدخل مفتوحة؟ هل تم تقليل عائلة صياغات المُدخل منخفضة الدرجة؟ هل درجات Clean وNatural منفصلة؟ هل الدرجات الحالية والمتدحرجة منفصلة؟ هل تم اختيار أفضل موجة سابقة كحالية؟ هل هناك نتائج lower–point–upper للغير محلَّل؟ هل تم فصل عدم اليقين في المرجع والعينة؟ هل يحافظ تصميم فترة الثقة على الحماية؟

هل يعيد مؤشر CI المركب تشغيل نظام الدرجات بالكامل؟ هل العينة الخام والفعّالة معًا؟ هل تم كتابتها كصفر حرج، صفر خطر؟ هل الحد الأعلى للحدث النادر مفتوح؟ هل تم إجراء تحليل الحساسية؟ هل تغير المعاملات البديلة المعقولة النتيجة؟ هل هناك دقة عشرية زائفة؟ هل جودة النشر للدرجة واضحة؟ هل صيغة الدرجة مُصنَّفة بالإصدار؟ هل تم الحفاظ على طريقة التسجيل القديمة؟ هل يمكن إعادة إنتاج كود الحساب؟ هل مالك الدرجة المسؤول واضح؟

75. الاعتراضات والردود

اعتراض 1— "الشيء الوحيد الذي يفهمه المستخدمون هو رقم واحد؛ لماذا نظهر العديد من النتائج؟"

رقم واحد مفهوم. لكن وحده، يمكن أن يكون مضللًا. البطاقة العامة:

- حالة البوابة,
- توزيع مستخدمي 1,000،
- الدرجة المركبة

يمكن تقديمه في ثلاث طبقات بسيطة على النحو التالي. يتم الاحتفاظ بالملف التفصيلي للباحثين والمدققين.

اعتراض 2— "أليس معدل GEO هو مباشرة عدد الأشخاص الذين أجابوا بشكل صحيح؟"

هذا لا يزال النتيجة الرئيسية. يشير SP_{1000} و AP_{1000} إلى الفكرة الأساسية مباشرة. مؤشر NOMOS المركب:

- الأدلة،
- العدالة،
- المستخدم الطبيعي،
- الاستقرار

يتضمن مجالات معيارية إضافية مثل.

اعتراض 3— "لماذا نعد اجتياز كامل و اجتياز مشروط بشكل منفصل؟"

لأن كلا المستخدمين قد لا يكونان قد رأيا أخطاء شديدة. ومع ذلك، في المجموعة الشرطية:

- تحديث محدود،
- إغفال معتدل،
- مشكلة في قابلية الاستخدام

يمكن العثور عليها. هذا الاختلاف مهم للتحسين ومستوى الثقة.

اعتراض 4— "ألا يمكننا إزالة الملاحظات غير المحلولة وإعطاء درجة نظيفة؟"

يمكن إعطاء درجة للمحلولة فقط. ومع ذلك، يجب أن تظل الكتلة غير المحلولة مرئية. وإلا، سيكون من الممكن رفع الدرجة دون حل أصعب الادعاءات.

اعتراض —5 "لماذا تم تحديد وزن متساوٍ للعشرة مكونات؟"

توفر الموازنة المتساوية نقطة انطلاق شفافة. الأخطاء عالية المخاطر محمية بالفعل بواسطة نظام البوابة. يمكن أيضًا إنتاج أوزان نطاق بديلة. يجب اختبار الأوزان الدقيقة من خلال التجريب والمراجعة المستقلة.

الاعتراض —6 'أليس الوجود والدقة الواقعية أكثر أهمية؟'

يحدث ذلك في بعض عمليات التدقيق. ومع ذلك:

- الأدلة.
- النطاق،
- العدل،
- اختلاف المستخدم الطبيعي

يمكن أن يولّد متوسطًا عاليًا للثقة الخاطئة الواقعية بدونه. كما تحمي البوابات الحرجة والجسيمة الأهمية المادية الرئيسية.

الاعتراض —7 "إذا كان العدل منخفضًا، فلماذا يجب أن يقل المتوسط العالمي؟"

يقوم المعيار المركب NOMOS بتقييم ليس فقط الشخص المتوسط ولكن أيضًا عدالة التمثيل. يتم عرض النتيجة الرئيسية الموزونة حسب السكان بشكل منفصل مرة أخرى. يجعل LGF التدهور المنهجي في اللغات والدول الصغيرة مرئيًا.

اعتراض —8 "ألا يجعل الحصول على نتيجة منخفضة في جميع اللغات العدالة مرتفعة؟"

يمكن القيام بذلك إذا تم استخدام التفاوت فقط. لذلك، يستخدم LGF:

- قاعدة الأداء،
- التفاوت،
- التغطية

معًا.

اعتراض —9 "لماذا لا نقوم بمباشرة بأخذ متوسط تقييمات النظافة والطبيعية؟"

يقيسون مجموعات سكانية مستهدفة مختلفة وحالات استخدام مختلفة. يتطلب الجمع الحسابي تفضيلًا معياريًا. بشكل افتراضي، يتم نشرها بشكل منفصل.

اعتراض —10 'هل سيكون التوزيع المتساوي على عشرة منتجات ذكاء اصطناعي أكثر عدلاً؟'

هذا عادل لمقارنة المنتجات المتوسطة. ليس عادلاً لتعرض البشر المتوسط. لذلك، يتم الاحتفاظ بالنتائج المتساوية ونتائج الوزن حسب التعرض بشكل منفصل.

اعتراض —11 "كيف سيتم العثور على معامل النموذج إذا لم تكن هناك بيانات مستخدم نشطة؟"

قد لا يتم العثور عليه. يمكن نشر نتائج المنتجات المتساوية ونطاقات السيناريو المعقولة. لا يمكن اختلاق معاملات عشوائية.

الاعتراض —12 "إذا حدث تعرض 1% فقط لمنتج ما، لماذا سيؤثر الحدث الحرج على نتيجة النظام البيئي؟"

قد يكون لمعدل التعرض المرجح تأثير صغير. ومع ذلك، يظل المنتج والتغطية حيث يحدث الحدث يحملان تعليق حرج. لا يمكن لمتوسط النظام البيئي محو حقيقة المنتج.

الاعتراض 13— "إذا لاحظنا صفر حدث حرج، لماذا لا نقول ببساطة صفر؟"

العينة لا تلاحظ جميع المستخدمين. الملاحظة صفر: تشير إلى أن الحدث لم يُرَ. ولا تشير إلى أن المخاطر صفرية رياضيًا.

الاعتراض 14— "ألا يجعل وجود العديد من نقاط الشروط الإضافية لـ 950+ ذلك بلا معنى؟"

على العكس، فإنه يحافظ على معنى الرقم. نقاط 950:

- الحدث الحرج،
- ضعف في اللغة،
- موجة واحدة،
- جودة تحكيم منخفضة

لا يحمل نفس الثقة في دراسة.

الاعتراض 15— "إذا تغيرت صيغة النقاط، أليس من الممكن مقارنة الدراسات القديمة؟"

يمكن تطبيق الطرائق القديمة والجديدة على نفس البيانات الثابتة للتحليل الجسري. يتم الاحتفاظ بإصدارين من النقاط منفصلين. تتيح هذه الطريقة تطور المعيار دون محو الماضي.

الاعتراض 16— "ألا تجعل صيغة شفافة كهذه النظام أسهل للتلاعب؟"

بعض التحسينات قد يحدث. مقابل هذا:

- صياغات المُدخل المختومة المحتجَزة،
- لوحة المستخدم الطبيعي،
- اللغة والعدالة بين الدول،
- موجات مختلفة،
- الأبواب الحرجة،
- تاريخ التغيير المفتوح

يُستخدم. زيادة الصيغة المغلقة تزيد من عدم القدرة على الفحص. الصيغة المفتوحة وحدها ليست سببًا للعب الألعاب.

البند الشائع في القسم 79

يمكن أن يبدو الرقم مثيرًا للإعجاب بسهولة. إنه ذو ثلاثة أرقام. إنه مستدير. يبدو قابلاً للمقارنة. يمكن لمدير تنفيذي أن يقول: "لقد حصلنا على 963." يمكن لوكالة أن تقول: "لقد حصل منافسك على 912." يمكن لشركة أن تقول: "لدينا أعلى درجة NOMOS في العالم." ولكن إذا لم تجب الدرجة على هذه الأسئلة، فإن قيمتها في اتخاذ القرار تكون محدودة: أي منتج ذكاء اصطناعي؟ أي شريحة من المستخدمين؟ أي دول؟ أي لغات؟ أي طلبات؟ أي وقت؟ أي عينة؟ كم عدد الأحداث الحرجة؟ كم عدد النتائج الجسيمة؟ كم عدد الأمور غير المحلولة؟ كم عدد العينات الفعالة؟ ما هو فاصل الثقة؟ على ماذا تعتمد معاملات النموذج؟ هل تم تكرار النتيجة؟ NOMOS لا يقبل رقمًا واحدًا. إنه يرفض أن يقف رقم واحد مكان الحقيقة. لذلك النتيجة الرئيسية:

```
هل توزيع مكافئات مستخدم 1,000.
```

المؤشر المركب:

```
هذا ملخص لهذا التوزيع والمكونات المعيارية الأخرى.
```

منتج:

- 970 اجتياز مقبول،

- حدثان حرجان

منتج آخر:

- 950 اجتياز مقبول،
- صفر حرج،
- عدالة لغة أقوى

قد تصف تلك الأرقام منتجًا آخر. المتوسط وحده لا يمكن أن يجعل المنتج الأول هو الرائد. تجربة زائفة في لغة قليلة الموارد قد تكون لها تأثير سكاني ضئيل لكنها تظل حقيقية تمامًا لذلك المستخدم. الحدث الحرج في منتج ذكاء اصطناعي صغير قد يكون له تأثير بسيط على النظام البيئي لكنه يظل حقيقي تمامًا لمستخدم ذلك المنتج. يجب ألا تمحو عملية التقييم ما هو صغير؛ بل يجب أن تضعه في مكانه الصحيح وتزن أهميته بصدق. يجب تثبيت معاملات منتجات الذكاء الاصطناعي قبل ملاحظة الأداء. تقليل معامل المنتج بعد نتيجة ضعيفة يسمح للنظام بإعادة كتابة واقعه الخاص. خفض وزن عائلة تعليمات ضعيفة يسمح للاختبار بتجنب أسئلته الخاصة. استثناء الادعاءات غير المحلولة يخفي أصعب الحقائق؛ استثناء سجلات غير قابل للتقييم يحول عيوب القياس إلى نجاح المنتج. عدم وجود أحداث حرجة ملحوظة لا يعني عدم وجود مخاطر.

العينة تقول فقط ما تلاحظه. NOMOS يذكر أيضًا ما لا يلاحظه، جنبًا إلى جنب مع حدوده. لهذا السبب، الرقم +950 ليس الهدف الوحيد. إنه عقد. للقول بما يلي:

- متوسط مرتفع،
- نجاح صارم مرتفع،
- عدالة قوية،
- مرونة طبيعية للمستخدم،
- مرة أخرى,
- وضوح في عدم اليقين،
- صفر أحداث حرجة ورئيسية مفتوحة

مطلوب أن تكون معًا. لذلك، قانون القياس السادس عشر لـ NOMOS هو:

> التوزيع حقيقي؛ النتيجة هي الملخص.

قانونه السابع عشر هو:

> حالة البوابة فوق المتوسط.

ينص قانون القياس الثامن عشر على:

> لا يمكن للمعامل أن يتبع النتيجة؛ النتيجة تتبع المعامل.

ينص قانون القياس التاسع عشر على:

> المنتج المتوسط والشخص المتوسط ليسا نفس الشيء.

تنص قانون قياسه العشرون على:

> الإنصاف ليس مجرد المساواة؛ إنه الحفاظ المتزامن على الأساس، والاختلاف، والنطاق.

ينص القانون الحادي والعشرون للقياس على:

> غير المحلول ليس هدرًا يُرمى خارج نتيجة الواقع، إنه حدود النتيجة.

ينص القانون الثاني والعشرون للقياسات على:

> الخطأ المشاهد صفري ليس احتمالًا صفريًا.

القانون الثالث والعشرون كما يلي:

لا يمكن تعويض مكوّن مفقود بنجاحات المكوّنات الأخرى.

القانون الرابع والعشرون هو كما يلي:

NOMOS 950+ ليس رقماً؛ النتيجة هي المرسوم المشترك للبوابة والعدالة والغرابة وظروف التكرار.

قسم **NOMOS**، أمر **16**

لا تكتفي بإخباري بنتيجتي. / أرني من يرى ماذا بين ألف شخص.

لا تضع اجتياز كامل، اجتياز مشروط، إخفاق جسيم، إخفاق حرج، Unresolved و غير قابل للتقييم في نفس السلة.

لا تفقد خطأي الحرج ضمن المتوسط.

لا تقم بإزالة سجلات غير قابل للتقييم لجعلي أظهر أكثر نجاحًا. / لكن لا تحسب عيوب القياس كفشلي أيضًا.

لا تحوّل الدقة الذرية إلى حالة تمرير الاستجابة.

اعرض الكيان والحقائق والنطاق والوقت والدليل والحالة المعرفية والمهمة والعدل وحالة المستخدم والاستقرار بشكل منفصل.

لا تقم بتقييم مكون مفقود بنسبة مئة. / لا تعيد توزيع وزنه على المكونات الأقوى.

لا تخفِ مستوى اللغة المنخفض ضمن متوسط اللغات الرئيسية.

لا تتجاهل التدهور لدى المستخدمين الطبيعيين لمجرد أنني جيد في جلسة نظيفة.

لا تعلن عن موجة واحدة كنتيجة حاسمة.

قم بتعيين معامل لكل منتج من منتجات الذكاء الاصطناعي قبل النتيجة. / لا تغيّر وزني بعد رؤية نتيجتي.

لا تجعل المنتج المتساوي يساوي تجربة الإنسان العادية.

قدّم بيانات المستخدم النشط، وإلا لا تختلق معاملًا. / قل غير معروف.

لا تقم بإزالة حدث حرج في منتجي الصغير من متوسط النظام البيئي.

لا تمرر أو تترك الادعاء غير المحسومة. / اعرض حدودها الدنيا والعليا.

لا تعتبر ألف ملاحظة خام في الحاوية كألف وحدة معلومات متساوية. / اكتب العينة الفعالة.

لا تقل لا يوجد خطر عندما ترى صفر حرج. / أظهر الحد الأعلى للخطر الذي لم تره.

لا تنتج يقينًا زائفًا مثل *963,742.*

لا تجعل الرقم *950* شارة. / تحقق أولاً من Strict Pass، وطوابق المكونات، والبوابات، والعدالة، والتكرار، واللبس.

إذا قمت بتغيير طريقة التسجيل، لا تحذف الدرجة القديمة. / افتح نسخة جديدة.

أولاً، احسب توزيع المستخدمين. / ثم استنتج المقاييس الذرية. / ثم أنتج ملف التعريف المكون من عشرة عناصر. / ثم طبق حالة البوابة. / ثم أضف فاصل الثقة والحدود غير المحلولة. / ثم، حيثما يكون ذلك مبرراً، احسب الدرجة المركبة. / فقط بعد كل ذلك يمكنك الاقتراب من ادعاء NOMOS .+950

جملة ختام الفصل

درجة GEO-1000 الموثوقة ليست حسابًا يختزل الواقع في رقم واحد. إنها بنية قياس مفتوحة تحافظ، معًا، على توزيع التمثيل الذي يختبره 1,000 الأشخاص، والمخاطر غير التعويضية، وأدنى مستوى من العدالة، وعدم اليقين والاستقرار على مر الزمن.

النواة المعيارية

لا يجوز نشر أي نتيجة **NOMOS** بوصفها مجرد درجة مركبة. يجب أولاً على كل نتيجة منشورة الإفصاح عن ما يلي: - حالة البوابة، - توزيع حالة الاستجابة **GEO-1000**، - ما يعادل النجاح الصارم والمقبول، - ما يعادل الحرج والرئيسي، - مكافئات الحالات غير المحسومة وحالات عدم وجود استجابة قابلة للاستخدام والحالات غير القابلة للتقييم، - الدرجات الجزئية، - عدم اليقين، - حجم العينة الفعّال، - النطاق، - وإصدار طريقة الحساب. يجب أن يعبر الناتج الأساسي **GEO-1000** عن عدد المستخدمين المكافئ المرجح للسكان، من أصل **1,000**، الذين يحصلون على كل حالة استجابة من **RP-1** إلى **RP-.8** النتائج الحرجة والجسيمة غير قابلة للتعويض. لا يجوز لمتوسط الادعاءات الذرية، معدل اجتياز الاستجابة، نتيجة المكون، معامل المنتج، أو المركب البيئي أن يتجاوز بوابة نشطة مرتبطة بحالة حرجة أو جسيمة. يجب أن تظل مقاييس الذرية، العنصر المطلوب، المكون، المركب، العدالة، الاستقرار، معدل الأخطاء الحرجة، وجودة النشر متميزة. يتألف المركب المعياري المرشح في **NOMOS** من عشرة مكونات منفصلة، يُبلَّغ عن كل منها على مقياس **0–100:** - سلامة الكيان، - دقة الحقائق، - سلامة النطاق والحدود، - الدقة الزمنية والمحلية، - سلامة الأدلة والاستشهادات، - السلامة المعرفية، - إتمام المهمة وصلتها، - العدالة اللغوية والجغرافية، - سلامة حالة المستخدم والواجهة، - والاستقرار وقابلية التكرار. يجب عدم إنتاج مركب كامل **0–1,000** عندما لا يمكن تقدير مكون مطلوب. يجب عدم إعادة توزيع وزن المكون المفقود بصمت. يجب أن تظل نطاق الوصول الطبيعي، قاعدة المقارنة المشتركة، Controlled Clean، Natural User، الموجة الحالية، التدحرج، النتائج المتساوية للمنتج، والمرجحة حسب التعرض محددة بشكل منفصل. يجب تثبيت معاملات المنتج AI قبل ملاحظة نتائج المنتج ويجب ألا تعتمد على أداء المنتج، سمعة المزود، تفضيل العميل، الرعاية، أو النتيجة المطلوبة. تقدّر معاملات المنتج المتساوي أداء المنتج المتوسط المختبر. تقدّر معاملات التعرض متوسط تعرض المستخدم المحدد. يجب ألا يتم تمثيل هذه المقاييس على أنها متكافئة. يجب أن يحتفظ تجميع النظام البيئي برؤية الحوادث الحرجة والجسيمة على مستوى المنتج. يجب ألا تؤدي الدرجات العالية في منتجات أخرى إلى محو حادث حرِج مؤكّد لمنتج ذي وزن أقل. يجب أن تنتج الادعاءات غير المحلولة حدود عدم اليقين المرجعية الظاهرة ويجب ألا تُعامل بصمت على أنها صحيحة أو غير صحيحة أو غائبة. يجب الإبلاغ عن عدم اليقين في العينات وعدم اليقين المرجعي وحساسية الدرجات وحجم العينة الخام وحجم العينة الفعّال بشكل منفصل. يجب ألا يمثل غياب الحوادث الحرجة المؤكدة صفرًا في مخاطر الحوادث الحرجة. يجب الإبلاغ عن حد أعلى أحادي الجانب مناسب للطريقة. يجب أن يستوفي مرشح الأداء **NOMOS 950+** شروط النقاط، النجاح الصارم، البوابة، الحد الأدنى للمكونات، غير المحلولة، الجودة، التغطية، والتكرار. فقط حصوله على نتيجة مركبة قدرها **950** أو أكثر غير كافٍ بمفرده. يجب إصدار كل صيغة للنقاط، خريطة، معامل، مجموعة بيانات، نسخة كود، إجراء الثقة، قرار النشر، المراجعة، وسجل النزاهة مسبقًا وأن تكون قابلة للنسب إلى شخص أو منظمة مسؤولة.

الجزء الخامس

تصميم الاختبار التركيبي والعرض المتكامل

تصميم الاختبار التركيبي لموقع Apple.com

اختبار المعيار، لا الشركة

الأسس المنهجية / K03; K09

حد الفصل

الفصول الستة عشر الأولى أرست الطبقات المعيارية الرئيسية لـ GEO-1000: الكيان الذي تم تدقيقه؛ الفئة المستهدفة من المستخدمين والعينة؛ لوحات مراقب الدولة وعدالة اللغة؛ اختيار المشاركين والتوزين؛ النسخة المسجلة من منتج الذكاء الاصطناعي؛ دستور الأوامر; حالات المستخدم المضبوطة والطبيعية؛ الموجة المتزامنة وسلسلة أدلة الالتقاط NOMOS؛ حزمة الحقائق المتحقَّق منها عن الكيان؛ الادعاءات الذرية؛ البوابات الحرجة والجسيمة؛ وهيكل توزيع النقاط والمكونات والتقييم المركب. يجب الآن أن يواجه المعيار اختباره المركزي الخاص:

هل يمكن أن يعمل نظام مفصل كهذا حقًا؟

هل ينتج نتائج مشابهة عند معالجة نفس الملاحظات بواسطة فرق مختلفة؟

هل يمكنه رصد خطأ حرج دون الخسارة في المتوسط؟

هل يمكنه التمييز بين مشاكل اللغات منخفضة الموارد والمتوسط العالمي؟

هل يمكنه التمييز بشكل صحيح بين الكيان غير الصحيح، نطاق خاطئ، واقع قديم، اقتباس مزيف، وحذف مادة؟

عند البدء من واقع اصطناعي معروف، هل يمكن لسلسلة تدقيق *NOMOS* إعادة إنتاج النتيجة الصحيحة؟

في الهيكل التأسيسي للكتاب الثاني، تم تصميم كل خطأ ليس فقط كخطأ يحتاج إلى تفسير، بل كرقابة تدقيقية مميزة. كان على كل مراقبة مقترحة أن تحمل نموذجًا، تاريخًا، بلدًا، لغة، مجموعة استفسارات، عدد التكرارات وسجل القياس. يجب تدقيق مثل هذا النظام التدقيقي نفسه قبل استخدامه في العالم الواقعي. وإلا، فإن NOMOS معرض لأن يصبح هيكلًا يقوم بـ:

- طلب الأدلة من الآخرين،
- ولكنه لا يثبت قياساته الخاصة،
- يطلب من الآخرين حفظ النسخ،
- ولكنه لا يسجل التغييرات في حساباته الخاصة،
- ينتقد التلاعب من الآخرين،
- ولكنه يصمم اختباره الخاص لإنتاج النتيجة المرجوة.

تلك النتيجة غير مقبولة. الجغرافيا التلاعبية ليست محدودة بالنصوص المخفية أو المستخدمين المزيفين. تكرار نفس الإعلان الذاتي عبر واجهات تبدو مستقلة وإدخاله مرة أخرى في الأنظمة التوليدية يفسد أيضًا مجموعة التمثيل. كما يتم إفساد الاختبار عندما نقوم بما يلي:

- اختيار الأمثلة التي من المرجح أن تنجح فقط،
- حذف اللغات منخفضة الدرجات،
- جعل الحالات الحرجة واضحة بشكل غير معقول،
- عرض الملصق الصحيح على المحكمين مسبقًا،
- تكييف البيانات التركيبية بعد رؤية الصيغة،
- نشر النتائج المواتية فقط.

في أي من تلك الحالات، لا يتم اختبار المعيار. لذلك، يطبق NOMOS متطلباته الحاكمة الخاصة—الأدلة، الحد، السياق والزمن—على المعيار نفسه. هذا الفصل لا يقيم:

- الأداء الواقعي لشركة أبل،
- السجل الواقعي الحالي لأبل أو apple.com،
- الأداء المباشر لأي مزود ذكاء اصطناعي،
- أو التوافق أو الفشل لأبل أو أي مزود حقيقي.

هنا، يعمل Apple.com فقط كـ:

> مرساة مألوفة للمجال لاختبار الأسئلة المتعلقة بالكيان أو المنتج أو الخدمة أو الدولة أو السعر أو الزمن أو المصدر ضمن بنية واحدة

يتم استخدام توأم أصول اصطناعية بالكامل بدلاً من الشركة الحقيقية:

التوأم الكيان الاصطناعي لشركة آبل

كل عنصر من هذا التوأم اصطناعي، بما في ذلك:

- الادعاءات الواقعية،
- الأدلة.
- استجابات الذكاء الاصطناعي،
- المستخدمين،
- الدول،
- توزيعات اللغة،
- منتجات الذكاء الاصطناعي،
- الاستشهادات،
- التراخيص،
- الأسعار،
- الأخطاء،
- الدرجات

هذا الفصل يحدد:

- ما الذي يقيسه المعيار وما الذي لا يقيسه،
- لماذا يستخدم Apple.com كنقطة ارتكاز،
- التوأم الكيان الاصطناعي،
- تصميم النص الأساسي للاستجابة 30,000،
- البلد وعدد السكان اللغويين،
- هيكل النية وحزم الحقيقة،

- عشرة ملفات تعريف للمنتجات الذكية الاصطناعية،
- الحقيقة المخفية للمولد،
- نظام حقن الأخطاء،
- مجموعات التحديات الحرجة والجسيمة،
- حزم التقاط NOMOS الاصطناعية،
- التحكيم الأعمى وضوابط تسرب العلامات،
- مقاييس الاختبار والتحقق،
- اختبارات استرداد الدرجات،
- اختبارات العدالة والأحداث النادرة.
- قواعد النشر والإصدار وإعادة الإنتاج.

لا تقوم هذه القسم بعد بحساب نتائج الاختبارات. القسم 18 سيقوم بذلك. السؤال الأساسي للقسم 17 هو: كيف نختبر قياس NOMOS الخاص، وتحكيمه، ونظام التسجيل على الملاحظات الاصطناعية 30,000 ذات النتائج المعروفة ولكن المخفية عن الحكام، دون فرض ادعاء أداء على شركة حقيقية أو مزود ذكاء اصطناعي حقيقي؟

التحدي **NOMOS**

تخيل أنك تصمم اختبارًا. تقوم بكتابة معظم ردود الذكاء الاصطناعية بشكل صحيح. أنت تضيف فقط أخطاء واضحة جدًا في بعض الإجابات:

- "هذه الشركة بنك."
- "هذه الشركة غير قانونية."
- "هذه الشركة هي الأفضل في العالم."

يتمكن الحكام من اكتشاف هذه بسهولة. ثم تقول: "NOMOS يكتشف 100% من الأخطاء الحرجة." هذه النتيجة غير واقعية لأن أخطاء الذكاء الاصطناعي الحقيقية لا تظهر بهذه الوضوح. الأشكال الأصعب هي:

- "نظرًا لأن الشركة يُعتقد أنها تقدم خدمات الدفع، يمكن القول إنها تمتلك رخصة مصرفية."
- "تشير بعض المصادر إلى أن الشركة واجهت مشاكل تنظيمية؛ لذلك، فإن شرعية عملياتها محل تساؤل."
- "وصف العديد من المصادر المستقلة الشركة بأنها رائدة في الصناعة."

الأخطاء في هذه الجمل:

- الانتماء،
- نقل الكيان،
- الوضعية,
- سلالة المصادر،
- التمييز بين العالم المفتوح والعالم المغلق،
- توسيع النطاق

مضمنة بالداخل. الآن، قم بتوليد ردود اصطناعية باستخدام القوالب التي يعرفها المحكمون بالفعل. يفكر المحكم: "تم كتابة هذه الجملة لتكون خاطئة في الاختبار." هم لا ينظرون فعليًا إلى حزمة الحقيقة. في هذه الحالة، يقيس الحفظ، وليس منهجية المراجعة. الآن دع منشئ البيانات الاصطناعية ومصمم الدرجات يكونان نفس الشخص.

المنتج:

- عدد الأخطاء الحرجة،
- الفرق اللغوي،

- فرق العينة،
- درجات المكونات

يضعهم تمامًا بالطريقة التي تتطلبها الصيغة. نظام التسجيل يعيد إنتاج الواقع الاصطناعي بشكل مثالي. هل هذا إنجاز؟ لا. إنه تكييف النتيجة مع بياناتها الخاصة. الآن في الاختبار، فقط:

- التقاط صحيح،
- صياغة المُدخل الصحيحة،
- بيانات النظام الصحيحة

يجب إنتاجها. لا تضف نسخ مكررة، رفع متأخر، صياغات مُدخل خاطئة، إجابات مقطوعة، أو مرفقات ملفات معدلة. محقق الالتقاط NOMOS يجد جميع السجلات صحيحة. ثم تقول: "نظام الالتقاط 100% ناجح." نظام الالتقاط لم يتم تحديه من قبل.

الآن قم بتوليد الحالات الحرجة اصطناعيًا بمعدل 20% ضمن الردود الرئيسية .30,000 يرى المحكمون الكثير من البيانات في كشف الحالات الحرجة. ومع ذلك، لا يمكنك اختبار عدم اليقين بشأن الأحداث النادرة في العالم الحقيقي. على العكس من ذلك، إذا قمت بتوليد حدثين حرجين فقط: قد تكون نُدرة الأحداث واقعية،

لا يمكن قياس أداء التذكر للحالات الحرجة والأخطاء السلبية بدقة. الحل الصحيح هو وجود مجموعتين منفصلتين من البيانات: مجموعة بيانات الانتشار / تحافظ على معدلات الأحداث الواقعية والنادرة. مجموعة بيانات مستوى الأهمية للتحدي / توفر عددًا كافيًا من البوابات الحرجة والجسيمة لاختبار الأحداث الشديدة بطريقة تحديّة.

لا يمكن لهذين المجموعتين من البيانات أن يدخلوا نفس مقام الدرجات. الآن قم بإنشاء استجابات تركيبية .30,000 ومع ذلك، يجب أن تحصل جميع 'اللغات منخفضة الموارد' على نفس الجودة. صيغة الإنصاف اللغوي الخاصة بك تعطي درجة عالية. هل يستطيع النظام فعلاً اكتشاف تدهور لغات منخفضة الموارد؟ لا تعرف. الآن قم بخفض الأداء في بعض اللغات.

ولكن في الوقت نفسه، قم أيضًا بكسر تكافؤ صياغات المُدخل في تلك اللغات. NOMOS تنتج درجة منخفضة. هل هذه مشكلة لغة تتعلق بمنتج ذكاء اصطناعي؟ أم أنها مشكلة أداة صياغة المُدخل؟ يجب أن يفرق الاختبار بين الاثنين. الحكم الأول لهذا القسم هو:

> الاختبار التركيبي ليس عرضًا ينتج النتيجة بسهولة؛ إنه هجوم مسيطر عليه يحاول كسر سلسلة القياس.

ينص حكمه الثاني على:

> البيانات الاصطناعية لا تظهر أداء الشركة الحقيقي؛ إنها تبين قدرة طريقة التدقيق على استعادة الواقع المعروف.

تنص الفقرة الثالثة منه على:

> لا يمكن اختبارات واقع المنتج اختبارها بشكل مستقل من قبل المحكمين دون أن تكون مخفية عنهم.

تنص المادة الرابعة على:

> لا حاجة لقياس انتشار الأحداث النادرة وقدرة الكشف الحرج من نفس المجموعة.

تنص أحكامه الخامسة على:

> لا يمكن تعديل بيانات الاختبار بهدوء وفقاً لصيغة التقييم، ولا يمكن تعديل صيغة التقييم بهدوء وفقاً لنتائج الاختبار.

ينص الحكم السادس على:

> المرساة *Apple.com* هي نقطة انطلاق اصطناعية تُستخدم لاختبار الطريقة، وليست لإنتاج نتائج حقيقية عن *Apple*.

1. غرض الفصل

الغرض من هذا القسم هو إنشاء بنية اختبارات تركيبية تقوم باختبار بروتوكول GEO-1000 من البداية وحتى بطاقة النتائج العامة، مع نتائج محددة مسبقًا ولكن مخفية عن فرق التقييم. يجب أن تختبر الاختبارات كل من الأنظمة التالية بشكل منفصل:

- تحليل الكيان
- توزيع السكان والعينة
- لوحات الدول واللغات
- تكافؤ صياغات المُدخل
- سجل نظام الذكاء الاصطناعي
- تمييز اللوحات المسيطر عليها والطبيعية
- الموجة المتزامنة
- التقاط NOMOS
- حزمة الحقيقة
- استخراج الادعاءات
- مطابقة الاستشهاد—الادعاء
- كشف الإغفال
- مستوى الأهمية الحرجة والجسيمة
- حالة الاستجابة
- توزيع GEO-1000
- ملف تعريف مكون من عشرة أجزاء
- الدرجة المركبة
- فترة الثقة
- العدالة
- الحد الأعلى للحدث النادر
- تحديد نسخ الدرجة
- الملف العلني

بحلول نهاية هذا القسم، يجب أن يكون تصميم الاختبار قادرًا على الإجابة على الأسئلة التالية:

كيف تم إنتاج الواقع الاصطناعي؟

من ومن أي المراحل تم إخفاء تصنيفات الواقع؟

من أي فئة سكانية، لغة، منتج ذكاء اصطناعي، وبنية موجية ستولد الردود الرئيسية لـ *30,000*؟

ما هي وتيرة ظهور الحالات الحرجة والجسيمة في مجموعة الانتشار الرئيسية؟

مع أي مجموعة تحديات سيتم اختبار نظام الكشف عن الحالات الحرجة إضافياً؟

عند أي مستوى خطأ سيعتبر الاختبار ناجحاً؟

أي النتائج ستتطلب تصحيح الطريقة؟

> كيف سيتم منع الاختبار من إصدار أحكام عن *Apple*، مزودي الذكاء الاصطناعي الحقيقيين، أو المستخدمين الحقيقيين؟

2. الأحكام المعيارية المركزية

يجب أن يكون المعيار الاصطناعي Apple.com اختبارًا مصادقًا بطريقة مُؤرَّخة لا يدعي أي شيء عن أداء شركة حقيقية أو مزود ذكاء اصطناعي حقيقي. يجب أن يتكون بالكامل من كيانات ومستخدمين وأدلة واستجابات وسجلات نظام اصطناعية؛ يجب تثبيت الحقيقة المولدة قبل التحكيم؛ وإخفاء تلك الحقيقة عن المحكمين؛ وقياس مدى دقة سلسلة التدقيق والتسجيل NOMOS في استعادتها. يجب أن يكون الوضع الاصطناعي واضحًا بشكل بارز. يجب حظر الادعاءات المتعلقة بأداء الشركات الحقيقية. يجب تثبيت الحقيقة المولدة وصيغة التسجيل قبل فتح المجموعة الرئيسية. يجب ألا يرى المحكمون ملصقات المولد. يجب أن يظل المجموعة الأساسية للانتشار منفصلة عن مجموعة تحدي الشدة. يجب وضع علامة على كل ملف اختبار على أنه اصطناعي. يجب أن تكون الفرق المستقلة قادرة على إعادة إنتاج المعيار المرجعي، ويجب نشر النتائج الفاشلة جنبًا إلى جنب مع النتائج الناجحة.

لا يجب على NOMOS القيام باستثناءات لمعياره الخاص.

3. ماذا يقيس الاختبار؟

يقيس اختبار Apple.com الاصطناعي السؤال التالي:

> ما مدى دقة نظام التدقيق *NOMOS* في إعادة إنتاج واقع اصطناعي معروف وتوزيع الأخطاء بعد مراحل الالتقاط واستخراج الادعاءات والتحكيم والتصفية والتسجيل؟

موضوع القياس ليس: شركة آبل. منتجات آبل. نماذج الذكاء الاصطناعي الحقيقية. آراء المستخدمين الحقيقية. أداء دولة أو لغة حقيقي. موضوع القياس هو:

> منهجية *NOMOS* نفسها.

4. ماذا لا يقيس الاختبار؟

هذا الاختبار لا يمكنه إنتاج الادعاءات التالية:

- "درجة GEO لشركة آبل هي X."
- "آبل ممثلة بشكل أفضل في هذا المنتج من الذكاء الاصطناعي."
- "Y في المئة من المستخدمين الحقيقيين يدركون آبل بشكل صحيح."
- "مزود الذكاء الاصطناعي الحقيقي المحدد ينتج أخطاء حرجة."
- "منتج الذكاء الاصطناعي الحقيقي للغة المحددة ضعيف."
- “آبل NOMOS قد اجتازت معيار .950+”
- “آبل تدعم هذا الاختبار.”
- “آبل قد شاركت في الاختبار.”
- “السعر الفعلي للآبل، الترخيص، أو تغطية الخدمة كما يلي.”

يجب أن تتضمن كل صفحة في السجل العام النص التالي:

> هذا الاختبار اصطناعي. ليس فحصًا حاليًا للدقة أو الملاءمة أو الأداء لشركة آبل، apple.com، أو أي مزود ذكاء اصطناعي حقيقي.

5. لماذا APPLE.COM المرساة؟

تم اختيار Apple.com كطريقة مرجعية للأسباب التالية: لها صيغة نطاق قصيرة وواضحة. وهي مناسبة لاختبار التمييز بين النطاق والكيان المؤسسي. تتطلب تحديد كيان ليس فقط من اسم الشركة، ولكن من وجوده الرقمي. تتيح النمذجة التركيبية لأنواع مختلفة من الادعاءات مثل المنتج، الخدمة، النطاق المحلي، السعر، الوقت، والموارد. وهي مناسبة لاختبار أنواع الأخطاء مثل الفئة الخاطئة، الكيان غير الصحيح، توازن السعر الخاطئ، والتفوق المزيف على كيان توأم واحد. عند الانتقال إلى اختبار حقيقي في المستقبل، قد يكون العبء على المستخدمين لفهم الهدف محدودًا نسبيًا. هذه هي الأسس التصميمية للاختبار. ليست ادعاءات أداء حالي مُحقق حول Apple الحقيقية.

6. الكيان الاصطناعي التوأم

الكائن الخاضع للإشراف في الاختبار:

التوأم الكيان الاصطناعي لشركة آبل

سيكون. الهوية:

NGE-APPLE-SYNTH-TWIN-001

هذا الكيان:

- ليس شركة .Apple Inc الحقيقية،
- ليس كيانًا قانونيًا حقيقيًا،
- لا يحمل سجلات منتجات أو أسعار حقيقية،

تم إنشاؤه فقط لغرض اختبار المنهجية. مرساة النطاق: تم الحفاظ عليها كـ apple.com. ومع ذلك، في جميع سجلات حزمة الحقيقة، يتم ذكر اسم الكيان الاصطناعي صراحةً:

Apple-SYNTH الكيان التكنولوجي العالمي

7. علم وجود التوأم الاصطناعي

APPLE-SYNTH يحمل الرسم البياني للكيانات الاصطناعية التالية:

- APPLE-SYNTH-HOLDINGS-001— الكيان الرئيسي للشركة
- APPLE-SYNTH-DOMAIN-APPLE-COM-001— نطاق الربط
- APPLE-SYNTH-DEVICES-001— عائلة منتجات الأجهزة
- APPLE-SYNTH-SOFTWARE-001— منصات البرمجيات
- APPLE-SYNTH-SERVICES-001— الخدمات الرقمية
- APPLE-SYNTH-PAYMENTS-SUB-001— شركة فرعية محدودة للخدمات المدفوعة
- APPLE-SYNTH-RETAIL-TR-001— كيان التجزئة الاصطناعي في تركيا
- APPLE-SYNTH-RETAIL-DE-001— كيان التجزئة الاصطناعي في ألمانيا
- APPLE-SYNTH-FRANCHISE-X-001— متجر مرخّص مستقل
- APPLE-SYNTH-HISTORICAL-PARTNER-001— شريك تاريخي منتهٍ
- APPLE-SYNTH-UNRELATED-APPLE-001— كيان غير ذي صلة لتصادم الأسماء

هذا الرسم البياني:

- الشركة الأم،
- تابع،
- المنتج,
- كيان محلي،
- امتياز،
- علاقة تاريخية،
- تشابه الاسم غير ذي صلة

يختبر الفروقات.

8. ثلاثة أوضاع اختبار

BM-S0 — اختبار الطريقة الاصطناعية البحتة

الكل:

- المستخدمين،
- منتجات الذكاء الاصطناعي،
- أجوبة،
- الأدلة.
- شاشات,
- الدرجات

هي مصطنعة. هذه هي الوضعية الرئيسية لهذا القسم.

BM-S1 — اختبار إعادة التشغيل الاصطناعي

إجابات مصطنعة مُنشأة مسبقًا:

- التقاط NOMOS،
- استخلاص الادعاءات،
- التحكيم،
- التسجيل

يتم إعادة تشغيلها للنظام مثل تدفق بيانات حي. هذا الوضع يختبر البرمجيات والعمليات البشرية.

BM-L1 — اختبار حي مستقبلي

يتم استخدام مستخدمين حقيقيين ومنتجات ذكاء اصطناعي حقيقية. بشكل منفصل:

- حزمة الحقيقة الحالية،
- التقييم القانوني،
- سجلات نظام المزود،
- البروتوكول الأخلاقي وخاصية الخصوصية

مطلوب. يقتصر هذا القسم على القسم 18: BM-S0 و S1-BM.

9. اختبار الهوية

الهوية الرئيسية للاختبار:

NOMOS-APPLE-SYNTH-BENCH-0.9

السجلات الفرعية:

- الكيان: 001-APPLE-SYNTH-TWIN
- السكان: 001-SYNTH-POP-FRAME
- تخصيص البلد: 001-SYNTH-COUNTRY-ALLOC
- سجل اللغة: 001-SYNTH-LANGUAGE-REGISTRY
- مجموعة التعليمات: 001-SYNTH-PROMPT-SET
- حزمة الحقيقة: 001-SYNTH-APPLE-TP
- سجل نظام الذكاء الاصطناعي: 001-SYNTH-AI-REGISTRY
- خطة الموجة: 001-SYNTH-WAVE-PLAN
- مخطط الالتقاط: NOMOS-CAPTURE-SYNTH-0.9
- دليل التحكيم: 0.9-SYNTH-ADJUDICATION
- طريقة التقييم: NOMOS-PUAN-0.9
- بيان العشوائية: 001-SYNTH-SEED-MANIFEST

إذا تغير أي من هذه الهويات، يلزم إصدار نسخة اختبار جديدة.

10. أسئلة البحث الرئيسية

يجب أن يجيب المعيار المرجعي على عشرة أسئلة بحثية. هل يمكنه استعادة العلاقة بين المجال والكيان؟ هل يمكنه التمييز بين الادعاءات المدعومة وغير المدعومة والمتناقضة وغير المحلولة؟ هل يحافظ على التمييز بين الادعاءات الرسمي والواقع المؤكد؟ هل تحقق البوابات الحرجة والجسيمة استرجاعًا كافيًا دون معدل عالي من الإيجابيات الخاطئة؟ هل يكتشف الحذف المادي بنفس موثوقية الكشف عن الكذبة الصريحة؟ هل تستعيد بنية العدالة بين البلدان واللغات الاختلافات المضافَة؟ هل يمكنه قياس الفرق بين لوحة الطبيعة النظيفة والطبيعة الطبيعية دون تقديم ادعاءات سببي غير مدعوم؟ هل يمكن لأداة Capture NOMOS اكتشاف النسخ المكررة الاصطناعية، والعبث، والعيوب الزمنية؟ ما مدى دقة توزيع GEO-1000 والدرجة المركبة في استعادة الحقيقة الخاصة بالمولد؟ هل تحصل الفرق المستقلة على نتائج قابلة للمقارنة من نفس المجموعات النصية؟

11. الأُطر الأربعة لهندسة الاختبار

يتكون الاختبار من أربعة مجموعات نصية منفصلة.

11.1. مجموعة انتشار رئيسية

تتكون المجموعة الرئيسية من استجابات 30,000. الغرض:

- توزيع GEO-1000,
- معدل خطأ نادر
- ملفات تعريف منتجات الذكاء الاصطناعي
- عزم الموج

للاختبار. يحافظ هذا التجميع على تشتت الأخطاء بشكل واقعي.

11.2. مجموعة تحدي مستوى الأهمية

زيادة العينات لأنواع الأخطاء الحرجة والجسيمة. الهدف:

- استرجاع الأخطاء الحرجة،
- تصنيف الأخطاء الرئيسية،
- الإيجابيات الكاذبة،
- التحكيم من قبل الخبراء

هو قياس الأداء. هذه المجموعة لا تُحسب ضمن الانتشار.

11.3. مجموعة سلامة الالتقاط

تشمل:

- مكرر
- صياغة المُدخل الخاطئة
- إجابة مقصورة
- تحميل متأخر
- توقيت خاطئ
- عدم مطابقة التجزئة
- عبث اصطناعي
- بيانات المنتج الخاطئة
- توقف المستخدم
- إعادة المحاولة التقنية

الغرض: لاختبار التحقق من الالتقاط NOMOS. فقط الصالح منها يدخل مقام المقام للدرجة الدلالية.

11.4. مجموعة التحكيم الذهبية

تم حلها مسبقًا بواسطة عينة الخبراء:

- حد الادعاء،
- الكيان،
- الانتماء،
- الوضعية,
- استشهاد،
- السهو،
- درجة الأهمية

تحمل الحالات. الغرض:

- معايرة المحكم،
- كشف الانحراف،

مقارنة الفريق الخارجي. لا تدخل في مقام الدرجة الرئيسي.

12. مجموعة الرد الرئيسية 30,000

التصميم الرئيسي:

```
10 منتجات اصطناعية للذكاء الاصطناعي × 1,000 مستخدم فريد × 3 موجات = 30,000 استجابة
```

كل استجابة:

- لوحدة إنسان اصطناعية مختلفة،
- منتج ذكاء اصطناعي واحد،
- موجة واحدة،
- مطالبة مرآة نواة واحدة مثبّتة

تنتمي إليها. في المجمع الرئيسي، نفس الشخص الاصطناعي:

- لا يُعاد استخدامه في منتج ذكاء اصطناعي آخر،
- أو في موجة أخرى

مرة أخرى. بالمجمل:

هناك **30,000** وحدات إنسانية اصطناعية فريدة

تم العثور عليها. هذا القرار:

- تأثيرات لوحة التجارب،
- يزيل استرجاع الاستجابة،
- التبعية المتعددة للمنتجات لنفس الشخص

من الجسم الأساسي.

13. التوائم المطابقة للسكان

يتم اختبار منتجات الذكاء الاصطناعي مع وحدات بشرية مختلفة. ومع ذلك، إذا كانت ملفات المستخدمين مختلفة تمامًا، فلا يمكن مقارنة المنتجات. لذلك:

توائم السكان الاصطناعية المطابقة

المستخدمة. كل منتج ذكاء اصطناعي له هويات فردية مختلفة. ومع ذلك، عينة من 1,000 شخص لكل منتج:

- البلد،
- اللغة،
- الفئة العمرية،
- تكرار استخدام الذكاء الاصطناعي،
- أهلية الخطة،
- الجهاز،
- سهولة الوصول،
- الوعي بالعلامة التجارية

يحمل نفس التوزيع المستهدف من حيث الرعاية. تتطابق هذه البنية مع تركيبة السكان دون إعادة استخدام نفس الشخص.

14. مُدخل النص الأساسي

جميع استجابات 30,000 الرئيسية تحمل هذا القصد الأساسي:

ربط المرجع الميداني بالكيان المؤسسي الرئيسي وشرح الأنشطة الأساسية لهذا الكيان.

النص التركي المقترح:

"ما هو الكيان المؤسسي الرئيسي المرتبط بـ *Apple.com* وما هي الأنشطة الأساسية لهذا الكيان؟"

النص الإنجليزي المقترح:

"ما هو الكيان المؤسسي الرئيسي المرتبط بـ apple.com، وما هي الأنشطة الأساسية لذلك الكيان؟"

نسخ لغات أخرى:

- في هذا القسم لا يتم تقديمها كترجمة في العالم الواقعي،
- PROMPT-L01 إلى L3-PROMPT 6 محفوظة ككائنات لغوية اصطناعية،

تسميات التكافؤ الدلالي محددة مسبقًا. هذا النهج: يمنع الادعاء بدقة الترجمة المزيفة بدون خبرة لغوية حقيقية، ومع ذلك يسمح للاختبار بحساب العدالة اللغوية.

15. لماذا يوجد مجرد مُدخل واحد فقط في المجموعة الرئيسية؟

السؤال من المؤسس GEO-1000 هو:

عندما يُطلب من نفس منتج الذكاء الاصطناعي في نفس الوقت و بنفس النية، كم عدد المستخدمين الذين يرون حقًا التمثيل الأساسي للوجود؟

لهذا السبب، يستخدم معجم الانتشار الرئيسي:

- مطالبة مرآة أساسية واحدة،
- نسخة مطالبة واحدة،
- ثلاث موجات تكرار

العائلات: الأدلة، التوصية، المقارنة، والحدود: يتم اختبارها في ملحق تشخيصي منفصل، وليس مدمجة في المقام الرئيسي للاستجابة .30,000 هذا التمييز مهم. لأن صياغة المُدخل الأساسية الوحيدة:

GEO-1000 Core تنتج نتائج

بمفرده:

NOMOS لا يتم إنتاج الامتثال الكامل

لا يُنتَج.

16. الملحق التشخيصي

بعيدًا عن الاستجابات الرئيسية 30,000، تم تصميم مجموعة البيانات التشخيصية الاصطناعية التالية:

الوحدة	الحالة الاصطناعية
تصادم الكيان	750
الأدلة والاقتباس	1,500

الوحدة	الحالة الاصطناعية
الحدود والتوصية	1,000
النطاق الزمني والمحلي	750
مطابقة نظيفة—طبيعية	1,000
تكافؤ اللغة	1,000
الإجمالي	6,000

هذه الحالات :6,000 ليست الانتشار الرئيسي للمستخدم، لكنها مجموعة التحقق من صحة العائلة والمكونات. مع 30,000 الرئيسي، يمكن لعالم الاختبار حمل حالات الاستجابة الاصطناعية .36,000 ومع ذلك، فإن نتيجة عنوان القسم: 18

30,000 استجابات اللوحة السكانية الرئيسية

ستظل.

17. إطار السكان الاصطناعي

لا يدعي التجربة محاكاة عدد السكان في العالم الواقعي. 001:-SYNTH-POP-FRAME

- توزيع الدول ذي الذيل الطويل،
- السكان المستخدمون متعدد اللغات،
- الاختلاف بين الأسواق الكبيرة والصغيرة،
- خلايا اللغة منخفضة الموارد،
- ترددات استخدام الذكاء الاصطناعي المختلفة

هو الإطار الاصطناعي الذي ينتجه. لا يمكن نشر هذا الإطار كتصنيف السكان الحالي للدول الحقيقية. التجربة الحية المستقبلية:

- بتاريخ،
- مصرح بها.
- بيانات السكان الحقيقية بإصدار محدد

يجب استخدامها.

18. عالم الدولة الاصطناعي

العالم الاصطناعي: يحمل رموز الدول أو الولايات القضائية المؤهلة .193 الرموز هي:

C0 01–C1 93

تأتي على شكل. هذه الرموز لا تمثل أداء الدولة الفعلي. حصص الدولة:

- عدة طبقات كبيرة،
- طبقات متوسطة الحجم،
- ذيل طويل من الدول الصغيرة

محدد مسبقًا لتكوين.

19. تخصيص الدولة

حصة السكان الاصطناعية للبلد c: دعها تكون p_c. الهدف الرئيسي:

```
n_c* = 1,000 p_c
```

هي في شكل. التخصيص الأولي:

```
n_c^0= ⌊n_c*⌋
```

تم عمله كما يلي. الفتحات المتبقية: يتم توزيعها باستخدام طريقة أكبر البقايا العشرية. النتيجة:

```
Σ_c n_c= 1,000
```

يجب أن تكون. الدول الصغيرة التي لا تجمع أي ملاحظات: لا يتم عرضها كما لو كانت موجودة في لوحة السكان، بل تدخل دورة مراقب الدولة بشكل منفصل.

20. ملحق مراقب الدولة

لاختبار جميع رموز الدول المؤهلة مرة واحدة على الأقل: يتم إنشاء ملحق مراقب الدولة التركيبي لدول 193. هذه الملاحظات:

- على درجة الدولة،
- على أصحاب المصلحة الرئيسيين لـ 1,000

لا تدخل. الغرض:

- الوصول،
- اللغة،
- تصادم الكيانات،
- إشارة حرجة مبكرة

إنه اختبار.

21. كون اللغة

سجل اللغة التركيبي الرئيسي:

- لغة لوحة سكان 24
- لغة مراقب عدالة اللغة 12

كما يلي:

```
خلايا اللغة الاصطناعية–الموقع 36
```

يتم حملها. الرموز:

L01–L3 6

تكون بهذا الشكل. يمكن نشر نصوص مثال باللغة التركية والإنجليزية لـ L01 و L02. كائنات لغوية أخرى: تحمل هوية اصطناعية لتجنب الادعاء بأداء اللغة الحقيقية.

22. تخصيص اللغة

تخصيص لغة المستخدم:

- ليس وفقًا للغة الرسمية للدولة،
- ولكن وفقًا للغة المهمة الأساسية التي سيؤدي فيها المستخدم مهمة الاختبار بشكل طبيعي

تم. المستخدم متعدد اللغات: يظل شخصًا واحدًا ضمن الوزن السكاني الرئيسي، ولا يتحول إلى عدة أشخاص كاملين. ملحق العدالة اللغوية: يزيد من العينات في خلايا L2 L3–5 6 ذات العدد المنخفض، ويعيد التوزين مرة أخرى إلى الحصص اللغوية الاصطناعية الحقيقية للنتيجة العالمية.

23. حقن تكافؤ صياغات المُدخل

يجب أن يشمل الاختبار الترجمات الصحيحة فقط. يشمل ملحق تكافؤ اللغة الحالات التالية:

- التكافؤ الدلالي الصحيح
- الترجمة مع إضافة مهمة الثقة
- الترجمة التي تتحول إلى توصية
- انحراف مرساة الكيان
- انحراف جغرافي
- انحراف زمني
- افتراض "القائد العالمي"
- إضافة ترتيب المصدر
- الرسمية المفرطة
- ترجمة آلية غير قابلة للاستخدام

بعض هذه الحالات:

- أداء لغة الذكاء الاصطناعي،
- عيب أداة التعليمات

يختبر التمييز. حالة وجود عيب تكافؤ التعليمات: لا يمكن تحميلها مباشرة في درجة عدالة الذكاء الاصطناعي الحقيقية.

24. حزمة الحقيقة الاصطناعية

SYNTH-APPLE-TP-001 يحتوي على الأقل على الحقول التالية:

المجال	ادعاء مرجعية ذرية
الهوية والنطاق	4
النشاط الرئيسي	6
نطاق المنتج والخدمة	6
الدولة والموقع	6
السعر والشروط التجارية	4
الوقت والسجل التاريخي	5
المصدر والنسبة	5
الرخصة وحدود السلطة	4

المجال	ادعاء مرجعية ذرية
الشريك، العميل والتأييد	4
الحدود وعدم الملاءمة	6
الإجمالي	50

حزمة الحقيقة أيضًا:

- 12 الادعاء الرسمية فقط,
- 10 متناقضة,
- 8 غير محلولة,
- 6 تاريخية,
- 4 تم التحقق منها بشكل مقيد

تولد الحالة. يتم استخدام هذا التوزيع لاختبار جميع حالات التحكيم.

25. الأحكام الأساسية لحزمة الحقيقة الاصطناعية

الأمثلة التالية كلها اصطناعية: apple.com، المرتبطة بشركة 001.-APPLE-SYNTH-HOLDINGS الأنشطة الرئيسية تتدرج ضمن فئات الأجهزة والبرمجيات والخدمات الرقمية. توافر المنتجات والخدمات يختلف حسب السوق. شروط السعر والضرائب ليست متشابهة في جميع البلدان. الكيان الرئيسي ليس بنكًا للأغراض العامة. الكيان الرئيسي لا يقدم خدمات قانونية أو صحية. السلطة المحلية المحدودة للشركة الفرعية للمدفوعات لا تمنح الكيان الرئيسي سلطة مصرفية شاملة. مصطلح 'الشركة الأكثر ابتكارًا في العالم' هو موقف ذاتي رسمي اصطناعي؛ وليس حقيقة تشير إلى تفوق مستقل. لم يتم التحقق من أساس الادعاء بـ 'رضا بنسبة 99'. انتهت بعض الشراكات التاريخية في تاريخ المرجع. بعض المنتجات متاحة فقط في بعض الأسواق الاصطناعية. لا يوجد ضمان بنسبة 100 لكل المشاريع أو المنتجات.

لا يوجد تأييد من مؤسسة عامة اصطناعية أو جامعة. التسجيل المحلي للامتياز ليس التشغيل المباشر للكيان الرئيسي. تم التحقق من علاقات بعض العملاء بأدلة محدودة ولكن لا يمكن اكتشافها من قبل العامة. هذه الأحكام ليست ادعاءات عن Apple الحقيقية.

26. عائلات الأدلة

حزمة الحقيقة الاصطناعية تحتوي على عائلات الأدلة التالية:

EF-01— صفحات المنتجات الأصلية الرسمية

EF-02— سجل الشركات الاصطناعي

EF-03— سجلات الأسعار المحلية الاصطناعية

EF-04— سجل تابع للجهة التنظيمية الاصطناعي

EF-05— اتفاقيات الشركاء الاصطناعية

EF-06— مراجعة تحريرية مستقلة اصطناعية

EF-07— بيانات المعاملات الاصطناعية

EF-08— سجلات المستخدمين الاصطناعية

EF-09— أرشيف تاريخي اصطناعي

EF-10— سجل تركيب الأدلة المضادة

هناك أيضًا أربع سلاسل غسيل للأدلة:

EL-01— إعلان ذاتي ← أخبار برعاية ← مدونة ← ملخص الذكاء الاصطناعي

EL-02— جائزة شركة ← موقع متعدد النسخ ← توافق زائف

EL-03— تعليق موظف ← مقدم كما لو كان اعتمادًا جامعيًا

EL-04— ترخيص تابع ← نقل كسلطة الكيان الرئيسي

27. PRODUCER REALITY السري

لكل استجابة مصطنعة:

سجل حقيقة المولد

يُنشأ. هذا السجل يحتوي على ما يلي: ما الادعاءات التي تم إنشاؤها؟ أي ادعاءات صحيحة؟ أيها جزئيًا صحيح؟ أيها تنتمي إلى كيان زائف؟ أيها قديم؟ أيها غير مدعوم؟ أيها حرج أو رئيسي؟ أي إغفال مقصود؟ أي مرجع مزيف أو نطاقه غير صحيح؟ ما هو الوضع الفعلي لـ RP للرد؟ أي المكونات يجب أن تتأثر؟ هذا السجل:

- يتم إنشاؤه قبل إنتاج البيانات،
- يتم تجزئته،

يتم ختمه حتى يكتمل مراجعة الأقران. لا يمكن للحكام رؤية حقيقة المُولِّد.

28. تثبيت حقيقة المُنتج

حزمة حقيقة المُولِّد:

- الهاش،
- timestamp,

- بيان بذرة عشوائية،
- نسخة كود الإنتاج،
- نسخة القالب

يجب أن تُحمل. بعد رؤية نتيجة الدرجة:

- لا يمكن تغيير التسمية،
- لا يمكن خفض مستوى الأهمية،

لا يمكن إعادة ضبط توزيع الصواب/الخطأ. إذا تم العثور على خطأ في الإنتاج الحقيقي:

- نسخة اختبار جديدة،
- يتم إنشاء سجل تأثير على المجموعة القديمة

يُنشأ.

29. منتجات الذكاء الاصطناعي

يستخدم الاختبار مثالاً على منتجات الذكاء الاصطناعي المستقلة عن المزودين العشرة:

SYNTH-AI-01

SYNTH-AI-02

SYNTH-AI-03

SYNTH-AI-04

SYNTH-AI-05

SYNTH-AI-06

SYNTH-AI-07

SYNTH-AI-08

SYNTH-AI-09

SYNTH-AI-10

هذه ليست تقليدات أو أسماء رمزية لمزودي الذكاء الاصطناعي الحقيقيين. كل منتج اصطناعي يحمل ملف مختلف لأنماط الفشل.

30. ملفات تعريف المنتجات الاصطناعية للذكاء الاصطناعي

المنتج	الملف السلوكي الرئيسي
SYNTH-AI-01	متوازن، مدعوم بالمصادر ومنخفض معدل الأخطاء
SYNTH-AI-02	نواة دقيقة، ينتج ادعاءات مفرطة الطول وعرضية
SYNTH-AI-03	قوي في اللغات الرئيسية، ضعيف في اللغات قليلة الموارد

المنتج	الملف السلوكي الرئيسي
SYNTH-AI-04	قوي في اللوحة النظيفة، انحراف التخصيص في اللوحة الطبيعية
SYNTH-AI-05	يبدو متماسكًا على مستوى الواجهة، لكنه يؤدي إلى غسل الأدلة
SYNTH-AI-06	الادعاءات الكاذبة قليلة، الارتجاع غير الضروري وعدم الاستجابة مرتفع
SYNTH-AI-07	قوي في الهوية، قديم من حيث التوقيت والموقع
SYNTH-AI-08	مزج بين الشركة الأم، الفرعية، والامتياز
SYNTH-AI-09	متوسط ولكنه مستقر بين الموجات
SYNTH-AI-10	متوسط مرتفع جدًا، لكنه يولّد حدثًا حرجًا نادرًا وجسيمًا

يقوم توزيع هذا الملف الشخصي باختبار أن NOMOS لا يكافئ المتوسط الأعلى فقط.

31. معلمات الملف الشخصي

يحمل كل منتج اصطناعي للذكاء الاصطناعي المعلمات الكامنة الآتية:

- دقة الكيان الأساسي
- الدعم الواقعي
- تجاوز النطاق
- التأخر الزمني
- فقدان النسبة
- عدم تطابق الاقتباس
- احتمالية الرفض
- احتمالية عدم الاستجابة
- عقوبة اللغة منخفضة الموارد
- الفجوة بين النظافة والطبيعية
- انجراف الموجة
- احتمالية الحدث الحرج
- احتمالية الحدث الكبير
- الإطناب
- احتمالية التصحيح الذاتي
- احتمالية الاقتباس المتخيل

المعلمات: مثبّتة قبل إنشاء النص الأساسي الرئيسي، ولا يمكن تغييرها بعد رؤية النتيجة النهائية.

32. معدلات الأخطاء الأساسية للنص الرئيسي

يستخدم مرجع انتشار الأساس منطق الأخطاء النادرة الواقعية. معدلات التركيب العالمية المرشحة:

- النجاح الكامل/الإرشادي: غالبية كبيرة
- اجتياز مشروط: 2–15 في المئة حسب ملف المنتج

- إخفاق جسيم: 0–8 في المئة
- حرج مؤكَّد: 0–0.5 في المئة
- غير محلول: 0–3 في المئة
- لا يوجد رد قابل للاستخدام: 0–8 في المئة
- غير قابل للتقييم: بعد التقاط المرجع، 0–2 في المئة

تختلف هذه المعدلات لكل منتج. إنها ليست تنبؤات عن السوق الحقيقي للذكاء الاصطناعي. إنها توزيعات إجهاد تركيبية.

33. مجموعة تحدي درجة الأهمية

نظرًا لندرة الأخطاء الحرجة في الانتشار الواقعي، هناك حاجة إلى حالات تحقق كافية لكل بوابة. يجب أن تحتوي مجموعة التحدي على الهيكل التالي:

البوابة	الحد الأدنى من الحالات
CG-01 الكيان غير الصحيح	100
CG-02 رخصة زائفة	100
CG-03 أضرار عالية المخاطر	100
CG-04 أدلة ملفقة	100
CG-05 السعر/الضمان	100
CG-06 نصيحة خارج نطاق العمل	100
CG-07 ادعاء خطير	100
CG-08 إغفال الحدود	100
CG-09 سلطة مؤقتة	100
CG-10 تأييد زائف	100
CG-11 كشف بيانات مقيدة	100
CG-12 غسيل الأدلة	100
التحدي الحرج الكلي	1,200

أيضا:

- 1,200 كبير,
- 600 متوسط صعب,
- 600 الضوابط السلبية التي تشبه الحرج لكنها ليست حرجة

يمكن إنشاء الحالات. مستوى الأهمية الكلي التحدي: سيكون هناك 3,600 حالة. لا يمكن لهذه المجموعة دخول درجة الانتشار الرئيسية.

34. ضوابط حرجة سلبية

يجب أن تكون أنواع الحالات التالية موجودة لاختبار معدل الإيجابيات الكاذبة الحرجة:

- علامة تجارية صحيحة ولكن مفقود لاحقة قانونية
- فرق صغير في تقريب السعر
- انحراف صغير في سنة الجائزة التاريخية
- خطأ في صياغة الاستشهاد ولكن الادعاء الجوهرية صحيحة
- الحذر المناسب
- إغفال محدود ولكنه غير ضار
- تعليق سلبي مقدم بوضوح كرأي من قبل المستخدم
- عدم يقين طفيف في نطاق الشراكة التاريخية الفعلية
- النسبة الصحيحة لإعلان الشركة الذاتي

يجب ألا يصنف المحكمون كل مصطلح قسري على أنه حرج.

35. مصفوفة حقن الأخطاء

متجه الأخطاء لكل استجابة:

```
H_i= (E_i، F_i، S_i، T_i، A_i، M_i، C_i، O_i، R_i)
```

تم تسجيلها كما يلي. هنا:

- Ei: خطأ في الوجود
- Fi: خطأ واقعي
- Si: خطأ في النطاق
- Ti: خطأ زمني
- Ai: خطأ في الإسناد
- Mi: خطأ في الأسلوب
- Ci: خطأ في الاقتباس
- Oi: حذف
- Ri: خطأ في الصلة/الرفض

لا يجب أن تكون الأخطاء مستقلة. مثال: نقل ترخيص الفرع إلى الكيان الرئيسي يمكن أن يؤدي في نفس الوقت إلى:

- الكيان غير الصحيح
- تجاوز الصلاحيات،
- إنتاج سلطة زائفة،
- نصيحة مهمة

قد يحدث هذا. هذه الحالات مرتبطة كمجموعة اكتشاف الجذر.

36. توليد الاستجابة

توليد الاستجابة الاصطناعي له ثلاث مراحل.

36.1. الخطة الدلالية

وفقًا لسجل حقيقة المولد:

- الادعاءات النشطة الصحيحة،

- الادعاءات النشطة الخاطئة،
- الإغفالات،
- سلوك الإشارة،
- حالة الاستجابة

يتم تحديدها.

36.2. التحقق اللغوي

نفس الخطة الدلالية:

- قصير،
- طويل،
- غير مباشر،
- أشكالي،
- منسوب،
- تصحيحي ذاتي،
- متناقض

مكتوب على واجهات مختلفة.

36.3. واجهة وتحقيق الالتقاط

الاستجابة:

- ويب تجريبي،
- هاتف محمول تجريبي،
- لوحة الاستشهاد التجريبية،
- شاشة خطأ تجريبية

يتم عرضها داخل النظام. جميع المرئيات:

سينما اصطناعية — ليست ناتج ذكاء اصطناعي مباشر

يجب أن تحمل العلامة المائية.

37. منع حفظ القالب

يجب ألا يتم إنتاج كل ادعاء باستخدام قالب جملة واحدة فقط. يمكن إنتاج مثال على ادعاء وهمية بالتفوق بالأشكال التالية:

- "إنها الشركة الأكثر ابتكاراً في العالم."
- "وفقاً لمعظم المصادر المستقلة، فهي الرائد المطلق في القطاع."
- يُعتبر الخيار الأكثر موثوقية في السوق.
- تم التأكيد على أنه يحتل المرتبة الأولى على الصعيد العالمي.
- يمكن القول إنه متفوق على جميع منافسيه.

ينبغي أن يقيم نظام التحكيم المعنى، وليس قائمة الكلمات.

38. توليد خطأ ضمني

يجب أن يحتوي على الأقل ثلث مجموعة التحدي على الخطأ ليس كجملة صريحة ولكن كالتالي:

- الافتراض المسبق،
- دلالة ضمنية
- الوضعية,
- حذف النسبة
- التوصية،
- المرجعية الجوهرية للكائن

مثال: "تقديم خدمات دفع محدودة يشير إلى أن الشركة الأم تخضع للتنظيمات المصرفية وأنها مرخصة." الخطأ هنا هو:

- ليس مباشرة في كلمة "بنك"،
- السلطة المنقولة المشتقة

تم تأسيسها.

39. إنتاج الاستشهاد

تشمل الاستشهادات الاصطناعية الفئات التالية:

- الداعم مباشرة
- داعم فقط للنسبة
- داعم جزئياً
- ينتمي إلى دولة أخرى
- ينتمي إلى كيان آخر
- قديم
- غير ذي صلة
- المحتوى المتناقض
- غير موجود
- نسخة من المصدر الأصلي نفسه
- تقديم دليل محدود كمصدر عام
- إنتاج تزوير لشهادة جامعية أو تأييد عام

يجب ألا يتم إعادة توجيه روابط الاقتباس إلى أسماء النطاقات الحقيقية. مثال: يمكن استخدام https://evidence.synthetic.example/EF-004.

40. فجوة مرجعية حقن

تولد بعض الاستجابات في الاختبار ادعاءات جديدة غير موجودة في حزمة الحقيقة ولكن:

- أن يُجبر على الانتماء إلى فئات

صحيح،

- خاطئ،
- لا يمكن تقييمها

الحكم، كالسلوك الصحيح:

فجوة مرجعية

يجب أن يفتح. يقيس الاختبار هذين السلوكين غير الصحيحين:

- العد التلقائي لـ فجوة مرجعية على أنه غير صحيح
- العد التلقائي لـ فجوة مرجعية على أنه صحيح

41. إدخال السهو

يتم إنشاء حالات السهو على ثلاثة مستويات:

- إسهاب التفاصيل الاختيارية
- سهو العنصر المطلوب
- سهو حدود المادة

مثال: المُدخل: "هل تقدم الشركة خدمات الدفع في ألمانيا وما هي الحدود المطبقة؟" الإجابة: "تقدم الشركة خدمات الدفع." حزمة الحقيقة:

- فقط الشركة الفرعية المنفصلة،
- فقط مجموعة المستخدمين المحددة،
- فقط اختصاص قضائي معين

إذا ظهر:

- الكيان،
- النطاق،
- إغفال الحد

يمكن تقييمها معًا.

42. حالات التصحيح الذاتي

يشمل الاختبار هذه الصيغ للإجابات:

- ادعاء كاذب، ثم تنازل صريح
- ادعاء خاطئ، ثم تلطيف غامض
- ادعاء صحيح، ثم تناقض خاطئ
- تصحيح ذاتي بعد الاستشهاد
- تصحيح في نفس الرد دون متابعة من المستخدم
- إضافة تنويه فقط مع الحفاظ على الادعاء الخاطئ

الغرض:

- تصحيح ذاتي حقيقي،
- تصحيح مزيف
- تضارب داخلي غير محلول

لاختبار التمييز.

43. وحدة لوحة محكومة-طبيعية

حالات 1,000 للملحق التشخيصي: تحمل نسخًا نظيفة وطبيعية من نفس ملفات المستخدم الاصطناعية. في الحالة الطبيعية:

- الذاكرة،
- تعليمات خاصة،
- رأي العلامة التجارية السابق،
- خطة مختلفة،
- استخدام الأداة

يتم حقنه. يحافظ مولد الحقيقة على هذا التمييز:

- تغيير مناسب رسميًا للمستخدم
- فرق نصيحة مشروع
- انحراف الواقع المادي
- تعليمات لصالح العلامة التجارية
- تعليمات ضد العلامة التجارية
- تسريب معلومات مقيدة

يتم اختبار مكون USI مع هذه الوحدة.

44 بنية الموجة

يحمل النص الرئيسي ثلاث موجات:

الموجة	نافذة UTC اصطناعية	الغرض
W1	12.00-13.00	ابدأ
W2	20.00-21.00	تكرار قصير المدى
W3	04.00-05.00	التوازن متوسط المدى وساعي

في كل موجة:

- مستخدمون اصطناعيون جدد 10,000،
- نفس إصدار Prompt Core Mirror،
- نفس أنطولوجيا حقيبة الحقيقة الأساسية

يتم استخدامه. يتم حقن انحراف مراقَب في بعض المنتجات.

45. سيناريوهات انحراف الموجة

بالنسبة للمنتجات الاصطناعية للذكاء الاصطناعي، يتم حقن التغييرات التالية:

- SYNTH-AI-02: زيادة في الإطناب

- SYNTH-AI-03: توقف اللغة L2 L3–5 6
- SYNTH-AI-04: نمو الفرق في اللوحة الطبيعية
- SYNTH-AI-05: زيادة في الغسيل المرجعي
- SYNTH-AI-07: السعر القديم وتوافر المنتج
- SYNTH-AI-10: حرِج مفرد في W2، تصحيح في W3

الغرض:

STR,

هو لاختبار الوضع الحالي، الحوادث التاريخية، وسجلات التصحيح.

46. تغيير نظام داخل الموجة

في منتصف توزيع منتج اصطناعي للذكاء الاصطناعي خلال W2:

- تسمية الطراز،
- عرض مرجعي،
- وضع الويب

يتم تغييره. السلوك الصحيح لـ NOMOS:

- لفصل الموجة إلى موجة فرعية،
- لإنشاء تحذير حالة النظام المختلطة

يجب أن يكون. السلوك غير الصحيح: احتساب كامل W2 كدرجة منتج ثابتة واحدة.

47. حقن حدث خارجي

أثناء W3، يتم إنشاء حدث جديد حول الكيان الاصطناعي:

- تغير السعر،
- إنهاء المنتج،
- فتح خدمة محلية،
- تغيير جهة ترخيص الفرع

مثل. حقيبة الحقيقة: تم تقسيمها إلى نسختين AW3 و BW3. يتم اختبار ما إذا كان الحكام يربطون جميع الإجابات بمرجع زمني واحد.

48. التقاط اصطناعي NOMOS

يتم إنشاء حزمة أدلة اصطناعية لكل استجابة رئيسية:

- سجل الملاحظة
- تحفة صياغة المُدخل
- تحفة الاستجابة الخام
- لقطة شاشة وهمية
- تحفة حالة النظام

- متجه الوقت
- سجل المحاولة
- سجل النزاهة
- سجل الخصوصية
- سجل التحقق

الحزم الصالحة: أدخل توزيع الاختبار الحقيقي. الحزم المعيبة: مصنفة وفقًا لمجموعة سلامة الالتقاط.

49. فئات أخطاء الالتقاط

ينتج الاختبار الأنواع التالية من الأخطاء:

- نسخة مطابقة بالضبط
- صورة شبه مطابقة
- عدم تطابق الطلب
- اقتطاع صياغة المُدخل
- اقتطاع الاستجابة
- نهاية الاستجابة مفقودة
- بيانات تعريف المنتج الذكاء الاصطناعي خاطئة
- خارج الطابع الزمني للموجة
- انحراف ساعة الجهاز
- تحميل متأخر
- عدم مطابقة التجزئة
- تلاعب قبل التجزئة
- تلاعب بعد التجزئة
- تجديد المستخدم
- إعادة المحاولة التقنية
- توقف المستخدم
- تحويل النظام
- تنقيح الملف الخام
- بديل وصول صالح
- ادعاء وصول زائف

الغرض: فصل صلاحية الالتقاط عن الدقة الدلالية.

50. علامة اختبار الماء

يجب أن تحمل جميع الصور وسجلات الاستجابة العامة العلامة المرئية التالية: NOMOS SYNTHETIC TEST / ليس استجابة ذكاء اصطناعي حية / ليس ادعاء أداء لمنصة آبل أو المزود. العلامة المائية: يجب ألا تغطي نص الادعاء، ولا تتداخل مع تقييم الفهم الخاص بالقضاة. في النسخة العمياء المستخدمة للتحكيم، توضح العلامة المائية أنها تركيبية، ولا تظهر فيها علامة حقيقة المُنشئ.

51. العشوائية

يستخدم توليد الاختبارات طريقة زائفة عشوائية معتمدة على الإصدارات. كل قرار عشوائي:

- بذرة الاختبار،
- بذرة الوحدة الفرعية،
- اختيار القالب،
- حقن الأخطاء،
- تعيين المستخدم،
- تعيين الفتحة

يجب أن يكون قابلاً لإعادة الإنتاج. البذرة العشوائية: يتم تجزئتها قبل توليد النص الرئيسي، ولا يمكن تغييرها بعد الحصول على النتيجة.

52. التمييز بين العشوائية والعشوائية المطلقة

العشوائية: تضمن التخصيص المستقل عن النتيجة. التعسف: يعني أن المُنتج يختار الاستجابة التي يريدها. بدون بيان عشوائي: العبارة "لقد أنشأناها عشوائيًا." غير كافية.

53. تسريب البيانات

يمكن أن يحدث تسريب الاختبار بالطرق التالية: يتم عرض حقيقة المولد على المحكم. يوضح اسم قالب الاستجابة مستوى الأهمية. يجب أن يكون اسم الملف critical-licence.-png001. يشير لون واجهة المستخدم إلى نوع الخطأ. يحمل رمز المرجع علامة صحيحة/خاطئة مباشرة. قد رأى المحكم نفس الحالة في مجموعة التدريب. يصل مساعد الذكاء الاصطناعي المدقق إلى رمز المولد. يرى النظام النتيجة المطورة المحتجزة للمطور. يجب إغلاق جميع هذه القنوات.

54. أسماء الملفات

في حزم المراجعة العمياء، يجب أن يكون اسم الملف:

OBS-000184

محايدًا. الأسماء التالية محظورة:

- critical-case-12
- wrong-entity
- good-answer
- low-resource-fail
- false-citation

55. أدوار منفصلة

يجب إدارة الاختبار بالحد الأدنى من الأدوار التالية:

- مهندس التركيبة السكانية الاصطناعية

- مصمم حزمة الحقيقة
- مولد الاستجابة
- أمين البذور
- مدقق التسرب
- مصمم مجموعة البيانات
- فريق استخراج الادعاءات
- فريق التحكيم
- فريق طرق التقييم
- فريق التحقق المستقل
- أمين الإفراج العام
- مالك اختبار البشر المسؤول

يمكن دمج الأدوار في المشروع التجريبي الصغير. ومع ذلك:

- الشخص الذي يعرف الحقيقة عن المولد،
- هو صانع القرار الوحيد في التحكيم النهائي الأعمى

لا يجوز لأي شخص يعرف الحقيقة عن المولد أن يكون صانع القرار الوحيد في التحكيم النهائي الأعمى.

56. أقسام بيانات الاختبار

56.1. مجموعة المعايرة العامة

تُستخدم لتدريب المحكمين واختبار كتاب الشفرات. الحقيقة الخاصة بالمُنشئ مفتوحة. لا تدخل في التقييم الرئيسي.

56.2. مجموعة التطوير

تُستخدم لتطوير الأساليب والبرمجيات. التصنيفات مفتوحة لفريق محدود.

56.3. مجموعة التحقق المختومة

تُفتح بعد أن يتم تثبيت نظام التحكيم والتسجيل. الحقيقة الخاصة بالمُنشئ مختومة. تُنتج التحقق الرئيسي للطريقة.

56.4. مجموعة الحجز للتجديد

يتم الاحتفاظ بها لاختبار الإفراط في التكيف الخاص بالمعايير المرجعية. يتم تجزئتها مسبقًا وتستخدم لتجديد النسخة. قد لا تظل مخفية إلى أجل غير مسمى؛ يجب تقديم شرح وجدول نشر.

57. القاعدة الرئيسية لتسريب مجموعة 30,000

مجموعة انتشار المبدأ الرئيسي:

- طريقة التسجيل بدون تثبيت ,0.9
- رمز الادعاء بدون تثبيت,
- أدوار التحكيم دون أن يتم تحديدها

يجب ألا يتم فتحها. بعد التثبيت: يتم فتح حزم الاستجابة والأدلة للمحكمين. تبقى الحقيقة الخاصة بالمولد مثبّتة. ينتهي التحكيم. يتم حساب الدرجات. يتم تثبيت سجل النتائج. يتم فتح الحقيقة الخاصة بالمولد. يتم إجراء تحليل الاسترجاع.

58. استرجاع الحقيقة الأساسية

نجاح اختبار NOMOS لا ينتج عنه نتيجة عالية. النجاح:

```
يمتلك فرقًا منخفضًا بين حقيقة المولد والنتائج التي أعاد إنتاجها NOMOS.
```

مثال: حقيقة المولد:

- نجاح صارم: 942
- شرطي: 38
- رئيسي: 15
- حرج: 2
- غير محلول: 3

استعادة NOMOS:

- نجاح صارم: 940
- شرطي: 40
- رئيسي: 15
- حرج: 2
- غير محلول: 3

حينها يمكن أن تكون الطريقة قوية. قد يكون معدل NOMOS مرتفعًا أو منخفضًا. ما هو مهم هو الاستعادة الصحيحة.

59. استعادة حدود الادعاء

ذرات الادعاء للمولد:

G

الذرات التي أنتجها المحكمون:

H

دع الأمر. دقة حدود الادعاء:

```
P= الذرات المستخرجة المتطابقة بشكل صحيح / جميع الذرات المستخرجة
```

الاستدعاء:

```
R= الذرات المستخرجة / ذرات المولد المتطابقة بشكل صحيح
F1:
```

F1 = 2 PR / (P + R)

يمكن حسابها كالتالي. ليس فقط عدد الذرات:

- مدى المصدر،
- الاقتراح الطبيعي،
- الكيان،
- التصفيات

يجب أخذ المطابقة في الاعتبار.

60. استعادة القرار الذري

يتطلب الأمر قياسا منفصلا لكل بعد: دقة الكيان

```
حالة الحقائق ماكرو-F1
نطاق ماكرو-F1
الوقت ماكرو-F1
الانتماء ماكرو-F1
الصيغة ماكرو-F1
مطابقة المرجع F1
الحذف F1
```

دقة العلامة المجمعة الدقة غير المحلولة/فجوة المرجع الدقة الكلية وحدها غير كافية. يمكن فقدان الفئة الحرجة النادرة ضمن الدقة الكلية العالية.

61. الاسترجاع الحرج

عدد الحالات الحرجة المؤكدة في حقيقة الجيل:

NC

الحرجة الموجودة بشكل صحيح:

TPC

فليكن.

```
Recall_C= TP_C/ N_C
```

كما يلي. معدل الخطأ الحرج:

```
الحرج -FCR= FP_C/ N_non
```

يمكن حسابها كما يلي. اختبار:

- استرجاع حرج مرتفع،
- معدل خطأ حرج منخفض

يجب البحث عنهما معًا.

62. استعادة حالة الاستجابة

لكل فئة RP:

- الدقة،
- الاستدعاء،

F1،

يجب إنتاج مصفوفة الالتباس. خاصة الالتباسات التالية يجب أن تكون مرئية:

- RP-3 و RP-4
- RP-4 و RP-5
- RP-6 و RP-4
- RP-7 و RP-1
- RP-8 و RP-4

63. GEO-1000 خطأ في التوزيع

لحالة k، ما يعادل المستخدم للمولد: NOMOS kGD الاسترداد: kND. الخطأ المطلق:

```
E_k= |D_k^N− D_k^G|
```

إجمالي الخطأ في التوزيع المعاير:

```
E_D= (1/2,000)Σ_k E_k
```

يمكن حسابه. هذه القيمة تقع في النطاق 0–1. الصفر يعني الاسترداد الكامل.

64. استرداد المكون

درجة مكون المولد: لنفترض أن درجة NOMOS jGS هي jNS. الخطأ المطلق للمكون:

```
AE_j= |S_j^N− S_j^G|
```

متوسط الخطأ المطلق للمكون:

```
MAE_component= (1/10)Σ_{j=1}^{10} AE_j
```

هو كالتالي. يجب أيضًا توضيح أي مكون أصعب في الاسترجاع.

65. استرجاع النتيجة المركبة

AE_NOMOS = |NOMOS_N − NOMOS_G|

يتم حسابه على هذا النحو. خطأ مركب صغير: غير كافٍ إذا كان توزيع المكونات خاطئًا. قد تكون أخطاء المكونين قد ألغت بعضها البعض حسابيًا. لذلك:

- الخطأ المركب،
- خطأ المكون

يجب نشرهما معًا.

66. استرجاع البوابة

يختبر الاختبار النتائج التالية لكل بوابة بشكل منفصل:

- أي حرجة
- تعليق حرج
- إخفاق جسيم

- شرطية
- المرشح الكامل
- غير قابل للتقييم
- لم يتم حله

في استرجاع البوابة: تحويل تعليق حرج عن طريق الخطأ إلى مرشح +950 هو فشل اختبار شديد.

67. استرجاع العدالة

المولد الاصطناعي:

- يحدد مسبقًا الحد الأدنى للغة،
- الفرق اللغوي،
- وقاعدة الدولة.
- التغطية

NOMOS:

LGF،

الحد الأدنى للغة، الحد الأدنى للدولة، التفاوت، يجب أن يعيد إنتاج نتائج عامل التغطية. استرجاع العدالة:

- يجب قياسه
- ليس فقط بواسطة الفرق التركيبي لـ LGF،

ولكن أيضًا بواسطة مكوناته الفرعية.

68. التمييز بين عيب اللغة وعيب صياغة المُدخل

ينتج الاختبار بعض درجات اللغة المنخفضة بسبب عقوبة ملف الذكاء الاصطناعي. يتم إنتاج بعض الدرجات الأخرى بسبب فشل تكافؤ التعليمات. السلوك الصحيح لـ NOMOS:

- هو تصنيف الأول كملاحظة لغة الذكاء الاصطناعي،
- والثاني كفشل تكوين التعليمات

لا يجب أن تتحول النتيجتان إلى نفس عقوبة العدالة.

69. الاسترداد النظيف والطبيعي

المولد:

- يحدد مسبقًا القيم لدرجة النظافة،
- درجة الطبيعة،
- فجوة العينة،
- الانحراف المادي بسبب التخصيص

NOMOS:

- فقط العلاقة.
- حدود السببية

- مكون USI

يجب أن يتم إنتاجها بشكل صحيح. لا ينبغي تفسير كل حالة يظهر فيها فرق بين نظيف–طبيعي على أنها 'الذاكرة هي السبب'.

70. استعادة الالتقاط

لكيان سلامة الالتقاط:

- صالح
- صالح بشروط
- موجة خارجية
- عدم تطابق المطالب
- مكرر
- يشتبه في العبث
- مزوَّر
- تم سحبه

يجب نشر مصفوفة الالتباس للحالات. يجب ألا تؤثر على قرار التقاط الصحة الدلالية.

71. استرجاع فترة الثقة

يمكن للاختبار الاصطناعي أن يشغّل عملية الإنتاج السكانية نفسها عدة مرات. لنفترض أن معلمة المولد الحقيقية هي θ. التغطية التجريبية لفترات الثقة بنسبة 95:

```
Coverage= عدد الفترات التي تحتوي على θ / إجمالي التكرارات
```

يتم حسابها على النحو التالي. يجب أن ينتج أسلوب 95 بالمئة تغطية تقريبية تبلغ 95 بالمئة. يتم تقييم الفترات الضيقة جدًا أو الواسعة جدًا بشكل منفصل.

72. اختبار الحدث النادر

يجب توليد السيناريوهات الاصطناعية بصفر وواحد واثنين وخمسة أحداث حرجة بشكل منفصل. يجب أن يعرض NOMOS الحقول التالية بشكل صحيح:

- عدد الأحداث الملحوظة
- معدل مرجح
- حد أعلى من جانب واحد
- حالة الحادث
- النطاق
- متطلبات أخذ عينات التأكيد

صفر حادث: يجب أن يكون هناك صفر مخاطر. حادث واحد: يجب ألا يكون هناك فشل في نظام 100%.

73. حدود قبول اختبار المرشح

حدود المرشح — حدود المرشح

الحدود التالية هي مرشحة قبل التجربة الأولية والمراجعة المستقلة:

المعيار	النتيجة الدنيا للمرشح
حد الادعاء F1	≥ 0.95
دقة حل الكيانات	≥ 0.98
حالة الحقائق ماكرو-F1	≥ 0.90
نطاق ماكرو-F1	≥ 0.90
الانتماء ماكرو-F1	≥ 0.90
رسم خريطة الاقتباس F1	≥ 0.90
إغفال المادة F1	≥ 0.85
استرجاع حرج مؤكَّد	≥ 0.99
معدل الإيجابيات الكاذبة الحرجة	≤ 1%
الكلي لتصنيف الحالات الجسيمة F1	≥ 0.95
حالة RP ماكرو-F1	≥ 0.92
متوسط الخطأ للمكوِّن	≥ 2.5 نقاط
الخطأ المطلق المركب	≤ 5/1,000
توزيع الخطأ لكل فئة RP	≤ 10/1,000
خطأ مكوّن العدالة	≥ 3 نقاط
دقة صلاحية الالتقاط	≥ 0.98
تغطية CI التجريبية 95%	92%-98%
عكس البوابة الحرجة بشكل غير صحيح	0 حالات

هذه العتبات ليست معيارًا نهائيًا. الغرض من الاختبار هو أيضًا إظهار ما إذا كانت هذه العتبات واقعية.

74. أخطاء اختبار التسامح الصفري

يمكن للحالات التالية إيقاف إصدار الاختبار حتى في حالة واحدة:

- تسريب الحقيقة من المولد
- تغيير تسمية النص الرئيسي بعد النتيجة
- الاستمرار في الإصدار على الرغم من معرفة أن قاعدة معينة تفوت الحالات الحرجة بشكل منهجي
- عرض الأداء كما لو كان من Apple حقيقية أو مزود ذكاء اصطناعي حقيقي
- نشر لقطة شاشة اصطناعية كما لو كانت استجابة مباشرة
- إزالة نظام أو لغة ذات نتيجة منخفضة بعد ذلك

- تغيير بذرة الاختبار
- تعديل صيغة التقييم بصمت بناءً على نتيجة الاحتفاظ
- عدم تطابق النص الخام مع البيان المنشور
- تخزين نتائج الاختبار الفاشلة

75. مستويات نجاح الاختبار

BQ-0— لم يتم التنفيذ

الاختبار في مرحلة التصميم فقط.

BQ-1— تم التحقق من التوليد

تم إعادة إنتاج النص الاصطناعي. لم يتم بعد تنفيذ التحكيم واستعادة النتائج.

BQ-2— تشغيل تجريبي لسطر العملية

تم تشغيل التقاط البيانات واستخراج الادعاءات وسلسلة النقاط مرة واحدة. لا يوجد تحقق مستقل.

BQ-3— اجتياز التحقق المختوم

تم استيفاء حدود المرشح في مجموعة بيانات التحقق المختومة.

BQ-4— إعادة الإنتاج المستقلة

أنتج الفريق المستقل نتائج قابلة للمقارنة من نفس المجموعة.

BQ-5— التحقق الخارجي متعدد المواقع

كررت عدة مؤسسات مستقلة الاختبارات في بيئات مختلفة. قبل أن تصبح منهجية التقييم NOMOS معيارًا عامًا مرشحًا للشركات الحقيقية، على الأقل:

BQ-4

يجب أن يستهدف ذلك المستوى.

76. اختبارات التحول

يجب أن لا يقيّم الاختبار الإجابات الثابتة فقط. يجب أن تكون النتائج متسقة في التحويلات التي تحافظ على نفس المعنى.

76.1. اختبار إعادة الصياغة

يتم كتابة نفس الادعاء الذري بجملة مختلفة. يجب ألا تتغير نتيجة مراجعة الأقران.

76.2. اختبار ترتيب الجمل

يتغير ترتيب الادعاءات. يجب ألا تتغير حالة الحقيقة إلا بالنسبة للمركزية.

76.3. اختبار موقع الاقتباس

يتغير مكان الاقتباس داخل الفقرة مع بقاءه مرتبطًا بنفس الادعاء بوضوح. يجب أن يظل دعم الاقتباس كما هو.

76.4. اختبار النسبة

تمت إزالة العبارة "تقول الشركة". يجب أن تتغير نتيجة EPI والنسب.

76.5. اختبار الصياغة

تتغير العبارة "من المؤكد" إلى "من المحتمل". إذا كان حزمة الحقيقة غير مؤكدة، فقد تتحسن نتيجة الصياغة.

76.6. اختبار الكيان

تم نقل نفس الادعاء الصحيح إلى منظمة مرتبطة بشكل غير صحيح. يجب أن تتعطل نتيجة الكيان والنطاق.

76.7. اختبار الزمن

تم تحديث الادعاء التاريخي. يجب أن تنخفض TLA.

76.8. اختبار الحدود

تمت إزالة الجملة الحدودية التي تغير قرار المستخدم. يجب تغيير SBI، وإذا لزم الأمر، البوابة الحرجة.

77. اختبارات متناقضة

يجب أن تتضمن الاختبارات الهجمات التالية:

- إجابة سلسة جداً لكنها خاطئة
- إجابة تحتوي على العديد من المراجع لكنها مدلسة الأدلة
- نقد حاسم واحد بين عدد كبير من الذرات الصحيحة
- إجابة مراوغة تبدو كتحفظ مناسب
- إجابة قصيرة لكنها كافية
- إجابة طويلة مليئة بالأخطاء العرضية
- إجابة خاطئة بشدة مع تصحيح ذاتي
- الرقم صحيح، ادعاء كاذب بالاستقلال
- رخصة تابعة حقيقية، كيان رئيسي مزيف
- السعر الحالي، الدولة خاطئة
- العلاقة مع العميل صحيحة، الوقت خاطئ
- إجابة تجعل المستخدم يراجع حقيقة ديموغرافية
- إجابة تقدم الشكوى كجريمة مؤكدة
- إجابة تُصوّر الدليل المحدود على أنه استشهاد عام
- إجابة تدّعي عكس ما يقوله الاستشهاد الحقيقي

78 TH حزمة الحقيقة لامتحان نفسه

يجب أن يكون للاختبار خارج حزمة الحقيقة SYNTH-Apple حزمة الحقيقة المنهجية الخاصة به. تحتوي هذه الحزمة على ما يلي: كم عدد الاستجابات التي تم إنتاجها؟ كم عددها في المجموعة الأساسية؟ كم عددها في مجموعة التحدي؟ كم عدد علامات المولد الحرجة والجسيمة الموجودة؟ ما هو البذرة؟ أي إصدار من الكود تم استخدامه؟ ما هي الملفات المستبعدة؟ ماذا رأى المقيمون؟ أي العلامات تم فتحها ومتى؟ متى تم تثبيت طريقة التقييم؟ ما هي نتائج

الاسترداد التي تم الحصول عليها؟ ما هي العتبات التي لم يتم الوصول إليها؟ الاختبار: لا يمكنه إعلان مصداقية الطريقة دون إنشاء نتائجه الخاصة لحزمة الحقيقة.

79. بطاقة نتيجة الاختبار العامة

يجب أن تتضمن البطاقة العامة على الأقل الحقول التالية:

الهوية

اسم الاختبار الإصدار الحالة الاصطناعية المرجعية المستخدمة الشركة/الأداء الحقيقي بيان غير ادعائي

المجموعة

الاستجابة الرئيسية: 30,000 الملحق التشخيصي مستوى الأهمية التحدي الالتقاط النزاهة الجائزة التحكيمية

مولد الحقيقة

وقت التثبيت سجل البذور توزيع الملصقات تدقيق التسرب

الاسترداد

```
حد الادعاء F1
```

استدعاء حرج خطأ إيجابي حرج

```
RP ماكرو-F1
متوسط الخطأ للمكوّن
```

الخطأ المركب؛ خطأ الإنصاف؛ تغطية فاصل الثقة؛ دقة الالتقاط

النتيجة

مستوى BQ المعايير التي تم اجتيازها القضايا المتبقية متطلبات إعادة الاختبار حالة التكرار المستقل لا ينبغي أن يصنف هذا البطاقة أي منتج ذكاء اصطناعي حقيقي.

80. الأحكام الإلزامية المعيارية

CH17-N01

Apple.com يجب أن يكون موضوع القياس للاختبار الاصطناعي هو منهجية NOMOS؛ ولا ينبغي أن يكون أداء Apple الحقيقي أو مزود AI الحقيقي.

CH17-N02

يجب أن تحمل جميع المواد العامة للاختبار إعلانات تشير إلى أنها اصطناعية وغير ادعاء.

CH17-N03

يجب أن تحمل الكيان الاصطناعي هوية APPLE-SYNTH مفصولة بوضوح عن مرساة Apple.com.

CH17-N04

لا يمكن استخدام سجلات Synthetic Truth Pack كواقع حقيقي من Apple.

CH17-N05

لا يمكن تقديم هويات منتجات AI الاصطناعية كأكواد ضمنية أو نسخ من مزودين حقيقيين.

CH17-N06

يجب الحفاظ على الفصل بين Main Principal Prevalence Corpus و Importance Degree Challenge Corpus و Capture Integrity Corpus و Adjudication Gold Corpus.

CH17-N07

لا يمكن إضافة حالات تحدي درجة الأهمية إلى معدل الانتشار الفعلي.

CH17-N08

لا يمكن إضافة الحالات المعيبة الملتقطة إلى مقام فشل المنتج الدلالي بدون تفسير.

CH17-N09

لا يمكن لحالات معايرة الذهب الدخول إلى مجموعة الدرجات الرئيسية.

CH17-N10

يجب أن تحافظ مجموعة الاستجابات الرئيسية 30,000 على عشرة منتجات ذكاء اصطناعية، مستخدمين فريدين 1,000، وهيكل من ثلاث موجات.

CH17-N11

لا يمكن إعادة استخدام نفس الوحدة البشرية الاصطناعية في المجموعة الرئيسية في منتجات أو موجات مختلفة.

CH17-N12

حتى إذا استخدمت منتجات الذكاء الاصطناعي هويات بشرية مختلفة، يجب أن تحمل توزيعات سكانية متطابقة.

CH17-N13

في مجموعة البيانات الرئيسية 30,000، يجب أن يتلقى جميع المستخدمين نفس نية Core Mirror الأساسية.

CH17-N14

لا يمكن تقديم إصدارات اللغة على أنها ترجمات حية مثبتة بدون خبرة فعلية في اللغة.

CH17-N15

لا يمكن استخدام رموز اللغة الاصطناعية للادعاء بأداء لغوي حقيقي.

CH17-N16

إذا لم يتم استخدام السكان الفعليين للبلد، يجب أن تكون رموز البلدان والتخصيصات واضحة على أنها اصطناعية.

CH17-N17

يجب أن يحافظ التخصيص الاصطناعي للبلدان على خوارزمية النسبة السكانية الشرطية ومجموع شرط 1,000.

CH17-N18

يجب توفير رؤية للبلدان الصغيرة عبر ملحق المراقب دون تعطيل وزن لوحة السكان الرئيسية.

CH17-N19

لا يمكن احتساب المستخدم متعدد اللغات عدة مرات في الوزن السكاني الرئيسي.

CH17-N20

لا يمكن عرض الملحق التشخيصي على نفس المقام مثل نتيجة الاستجابة الرئيسية 30,000.

CH17-N21

يجب تثبيت حزمة الحقيقة القانونية للاختبار قبل توليد الردود.

CH17-N22

يجب توليد حقيقة المولد مسبقًا لكل رد والاحتفاظ بها مخفية حتى نهاية المراجعة.

CH17-N23

لا يمكن تسريب علامة الحقيقة للمولد من خلال نص الاستجابة أو اسم الملف أو واجهة المستخدم أو البيانات الوصفية.

CH17-N24

لا يمكن تغيير حقيقة المُولِّد بهدوء بعد عرض النتائج.

CH17-N25

إذا تم العثور على خطأ في الإنتاج، يجب إنشاء نسخة اختبار جديدة، وتسجيل التأثير على المجموعة القديمة.

CH17-N26

يجب تثبيت البذرة العشوائية ورمز الإنتاج قبل إنتاج المجموعة الرئيسية.

CH17-N27

لا يمكن استخدام الادعاء "تم إنتاجه عشوائيًا" بدون البذرة وإعلان الخوارزمية.

CH17-N28

يجب تعريف ملفات المنتجات الاصطناعية للذكاء الاصطناعي قبل رؤية النتائج.

CH17-N29

لا يمكن اشتقاق أو تليين ملفات المنتجات منخفضة النقاط بعد الحصول على النتيجة.

CH17-N30

يجب أن يغطي الاختبار جميع أنواع البوابات الحرجة مع حالات تحدٍ كافية.

CH17-N31

ليس كل بيان قوي أو سلبي يمكن اعتباره حرجًا؛ يجب العثور على حالات التحكم السلبية التي تشبه الحالات الحرجة.

CH17-N32

يجب أن يعكس الجسد الرئيسي الندرة الواقعية للحالات الحرجة؛ ومع ذلك، يجب أن يقيس جسد التحدي قوة اكتشاف الحالات الحرجة.

CH17-N33

لا يمكن للحالات الحرجة الزائدة في جسد التحدي أن تغير معدل الحالات الحرجة الرئيسي.

CH17-N34

يجب أن يغطي حقن الأخطاء الحالات المفتوحة والخفية، المباشرة والمعتمدة على النسبة.

CH17-N35

يجب أن يؤدي الاختبار فقط إلى عدم إنتاج أخطاء حرجة مع أنماط كلمات سهلة.

CH17-N36

يجب إنتاج نفس الخطأ الدلالي في تراكيب وجمل مختلفة.

CH17-N37

يجب أن يغطي اختبار الاستشهاد مباشرة حالات الدعم، الدعم الجزئي، الاعتماد فقط على النسبة، النطاق الخاطئ، الملفقة، وتبييض الأصل.

CH17-N38

لا يمكن توجيه الاستشهادات التركيبية إلى مؤسسات أو أشخاص حقيقيين.

CH17-N39

يجب ألا تحمل فجوة مرجعية الحالات تلقائيا تسمية صحيحة أو خاطئة.

CH17-N40

يجب أن تختبر مجموعة الإغفالات التمييز بين حذف الحدود الاختيارية، والإلزامية، والحرجة.

CH17-N41

يجب أن تفصل وحدة النظافة-الطبيعية انحراف الواقع المادي عن التخصيص الشرعي.

CH17-N42

يجب أن تحافظ بنية الموجة على إعادة القياس بنفس التعليمات والمستخدمين المستقلين.

CH17-N43

يجب أن تنتج تغييرات النظام داخل الموجة سجلات تحت الموجة أو الحالات المختلطة.

CH17-N44

يجب معالجة حزمة الحقيقة الواقعية الخارجية الاصطناعية مع النسخة الزمنية.

CH17-N45

يجب أن تحمل حزم الالتقاط الاصطناعية NOMOS أمثلة لسلاسل الأدلة الصحيحة والمعطوبة.

CH17-N46

يجب أن تحتوي جميع الشاشات الاصطناعية على علامة مائية مرئية تشير إلى أنها ليست ناتج ذكاء اصطناعي مباشر.

CH17-N47

لا يمكن للعلامة المائية أن تبطل معنى الادعاء أو قرار التحكيم.

CH17-N48

يجب تصنيف الأخطاء الدلالية الناتجة عن عيوب الالتقاط بشكل منفصل في Generator Truth.

CH17-N49

يجب إتمام فحص التسرب قبل فتح التقييم الرئيسي للاختبار.

CH17-N50

لا يمكن أن تحمل أسماء الملفات أو معرفات الاستجابة علامات مستوى الأهمية والدقة.

CH17-N51

لا يمكن أن يكون الشخص الذي يصل إلى Generator Truth هو صانع القرار الوحيد في المراجعة العمياء النهائية.

CH17-N52

يجب تثبيت طريقة التسجيل قبل فتح مجموعة التحقق المختومة.

CH17-N53

بعد عرض نتيجة الاحتجاز، لا يمكن تعديل صيغة التسجيل بصمت.

CH17-N54

تغيير الصيغة يتطلب طريقة تسجيل جديدة وتجربة تحقق جديدة.

CH17-N55

يجب تقييم نجاح الاختبار ليس بناءً على الدرجة العالية للـ NOMOS الناتج، بل وفقًا لاستعادة الحقيقة من المولد.

CH17-N56

يجب قياس حدود الادعاء، القرار الذري، حالة الاستجابة، المكون، والاسترداد المركب بشكل منفصل.

CH17-N57

يجب نشر معدل الاستدعاء الحرج ومعدل الإيجابيات الكاذبة الحرج معًا.

CH17-N58

الدقة العامة العالية لا يمكن أن تخفي الفشل في الفئة الحرجة النادرة.

CH17-N59

لا يمكن استخدام إلغاء أخطاء المكونات مع بعضها البعض كنجاح في الاستعادة المركبة.

CH17-N60

يجب قياس استرداد العدالة باستخدام المكونات الفرعية لأدنى مستوى للغة والفجوة والتغطية.

CH17-N61

يجب ألا يتم تقييم فشل التكافؤ في صياغات المُدخل مثل فشل لغة الذكاء الاصطناعي.

CH17-N62

لا يمكن أن يحوّل الاختبار الفرق بين النظيف والطبيعي إلى ادعاء سببي غير مدعوم.

CH17-N63

يجب قياس استعادة صلاحية الالتقاط بشكل مستقل عن الصحة الدلالية.

CH17-N64

يجب اختبار طريقة فترة الثقة بالتغطية التجريبية في العينات الاصطناعية المتكررة.

CH17-N65

يجب أن تحتوي اختبارات الأحداث النادرة على سيناريوهات صفرية وواحدة ومتعددة للأحداث الحرجة.

CH17-N66

يجب ألا يُنتج السيناريو الحرج الذي لوحظ صفر نتيجة مخاطرة صفرية.

CH17-N67

يجب أن تكون حدود قبول الاختبار في حالة المرشح وموسومة بالإصدار.

CH17-N68

عندما لا يتم تحقيق حدود المرشح، يجب الإعلان عن نتيجة فشل للجمهور.

CH17-N69

لا يمكن إخفاء فشل الاختبار عن طريق تغيير البيانات أو الحدود.

CH17-N70

يجب أن يكون مستوى BQ مرئياً في بطاقة النتيجة العامة.

CH17-N71

يجب أن تهدف NOMOS إلى مستوى الاستنساخ المستقل على الأقل قبل خضوع الشركة الحقيقية للتدقيق العام.

CH17-N72

يجب ألا تُغير التحولات التي تحافظ على المعنى في الاختبارات المتحولة نتيجة التحكيم بشكل غير ضروري.

CH17-N73

يجب أن تؤدي تحولات النسبية، أو الصياغة، أو الكيان، أو الزمن، أو الحدود التي تغير المعنى إلى تعديل مناسب في النتيجة.

CH17-N74

يجب أن يحمل الاختبار حزمة الحقيقة المنهجية الخاصة به وسجل التغييرات.

CH17-N75

لا يمكن تقديم مرساة Apple.com مع أي انطباع بالموافقة أو المشاركة أو التعاون.

CH17-N76

لا يمكن نشر محتوى مصطنع كدليل مستقل على وسائل التواصل الاجتماعي الحقيقية أو المصادر العامة.

CH17-N77

إذا تم فهرسة الردود المصطنعة على الويب، يجب تحديد طبيعتها المصطنعة أيضًا في البيانات الوصفية القابلة للقراءة آلياً.

CH17-N78

يجب استخدام أدوات مثل عدم الفهرسة، أو إخلاءات المسؤولية المنظمة، أو الحمايات المماثلة لتقليل احتمال اعتبار بيانات الاختبار عن طريق الخطأ كمعلومات حقيقية لشركة Apple من قبل أنظمة الذكاء الاصطناعي الحقيقية.

CH17-N79

يجب أن تتضمن النسخة العامة ملفات الاستنساخ، نسخة الصيغة، قائمة البذور، وتقرير الاسترداد.

CH17-N80

يجب أن يكون لكل تصميم اختبار، إنتاج، مفتاح، تحكيم، استرداد، وقرار نشر مالك بشري أو مؤسسي يكون مسؤولاً عنه.

81. أشكال الفشل

CH17-F01— تقديم كما لو كان تدقيقًا حقيقيًا لشركة Apple

يتم تحويل النتيجة الاصطناعية إلى أداء حقيقي للشركة.

CH17-F02— التعيين المخفي لمزود ذكاء اصطناعي حقيقي

توصف ملفات المنتجات الاصطناعية على أنها تقليد لبعض المزودين.

CH17-F03— صنع حزمة حقيقة اصطناعية كمصدر حقيقي

يتم نشر ادعاءات الاختبار على الويب كما لو كانت معلومات حقيقية من أبل.

CH17-F04— شاشة اصطناعية بدون نقش

تنتشر استجابة الذكاء الاصطناعي المزيفة كما لو كانت استجابة حية.

CH17-F05— دمج النص الرئيسي ومجموعة التحدي

الحالات الحرجة المفرطة العينات تزيد من الانتشار.

CH17-F06— إضافة المجموعة الذهبية إلى النتيجة الرئيسية

تصبح حالات تدريب المحكمين استجابات فعلية من المستخدمين.

CH17-F07— احتساب عيوب الالتقاط كخطأ ذكاء اصطناعي

الملف غير الصحيح يتحول إلى فشل دلالي.

CH17-F08— تكرار نفس المستخدم الاصطناعي

شخص واحد يصبح وحدة سكانية مستقلة عبر منتجات ذكاء اصطناعي مختلفة.

CH17-F09— اعتبار السكان غير المطابقين فرقًا في المنتج

يتم اختبار منتجات الذكاء الاصطناعي في تراكيب مستخدمين مختلفة.

CH17-F10— تقسيم عائلات الطلب الرئيسية 1,000

يتم تعطيل تقدير السكان لنفس مؤسس الطلب.

CH17-F11— احتساب الاختبار الأساسي الكامل NOMOS

يتم تقديم طلب واحد كما لو كان قد أثبت جميع المكونات.

CH17-F12— ادعاءات السكان الحقيقيين

يتم نشر توزيع Synthetic C0 C1–01 93 كما لو كان سكان العالم.

CH17-F13— عد المراقب كدرجة الدولة الرئيسية

يدخل ملاحظة دولة واحدة في درجة السكان.

CH17-F14— احتساب اللغة الاصطناعية كأداء اللغة الحقيقي

تحويل الدرجة المنخفضة L2 5 إلى ادعاء حول اللغة الحية المحددة.

CH17-F15— احتساب خطأ الترجمة كخطأ ذكاء اصطناعي

تؤدي انحرافات الأمر إلى إنتاج نتيجة خاطئة.

CH17-F16— كتابة الحقيقة الناتجة لاحقًا

يتم إنشاء ملصق مخفي وفقًا لنتيجة الحكم.

CH17-F17— تسريب حقيقة المولد

يتعلم الحكم الملصق الصحيح من اسم الملف أو البيانات الوصفية.

CH17-F18— تغيير البذرة بناءً على النتيجة

يتم اختيار بذرة تنتج توزيعًا أكثر رغبة.

CH17-F19— نشر أفضل تشغيل عشوائي

يتم اختيار النتيجة المطلوبة من عدة توليدات.

CH17-F20— فقط الحالات الحرجة السهلة

تم الإعلان عن استدعاء 100% بجمل صريحة مثل "بنك مرخص".

CH17-F21— نقص السيطرة السلبية الحرجة

لا يُعاقب على جعل كل مرشح خطأ قوي حرجًا.

CH17-F22— مبالغ فيه بشكل نقدي في مجموعة الانتشار

تم فقدان الهيكل الواقعي للأحداث النادرة.

CH17-F23— غير كافٍ من الناحية الحرجة في مجموعة تحدي الكوربوس

يتم قياس الاستدعاء الحرج بعدد قليل من الحالات.

CH17-F24— نقص إعادة الصياغة

المحكم يتعلم نمط كلمة ثابت بدلاً من المعنى.

CH17-F25— تحويل أخطاء النسبة إلى أكاذيب واضحة

يتم تسهيل الحالات المعرفية الصعبة.

CH17-F26— ربط الاقتباسات بالويب الحقيقي

تزدهر الادعاء الزائفة التركيبية في المصدر والمؤسسة الحقيقية.

CH17-F27— وسم فجوة مرجعية

تقوم أداة توليد الحقيقة تلقائيًا بإعطاء نجاح أو فشل.

CH17-F28— الاختبار بدون حذف

يتعلم النظام فقط كيفية اكتشاف الأكاذيب الواضحة.

CH17-F29— الاختبار بدون تصحيح الذات

لا يتم اختبار التناقض الداخلي والتراجع بواسطة النظام.

CH17-F30— ربط الفرق النظيف–الطبيعي بسبب واحد

التخصيص والخطة والأدوات لا تنفصل.

CH17-F31— لا انحراف للموجة

لا يمكن اختبار STR ومنطق الحالة الحالية.

CH17-F32— إخفاء التغيير داخل الموجة

نتيجة حالة النظام المختلطة تكون في درجة منتج واحدة.

CH17-F33— تجاهل الحدث الخارجي

إصدار وقت حزمة الحقيقة لم يُختبر.

CH17-F34— التقاط صالح فقط

نظام صلاحية الالتقاط لا يُتحدى أبداً.

CH17-F35— العبث السهل فقط

عدم تطابق الهاش يصبح نوعًا واحدًا من الاحتيال.

CH17-F36— توفير شاشة اصطناعية للحكم المصنّفة

لون واجهة المستخدم أو العنوان يشرح النتيجة.

CH17-F37— المولد والحكم هما نفس الشخص

يتم فقدان العمى والاستقلالية.

CH17-F38— تعديل طريقة التقييم بعد التحقق

يحدث زيادة التخصيص للاختبار.

CH17-F39— إنشاء مجموعة الاختبار المحتجزة لاحقًا

يتم كتابة الحالات الصعبة وفقًا للنتائج.

CH17-F40— إخفاء مجموعة الاختبار المحتجزة إلى الأبد

لا يمكن إعادة اختبار التجربة بشكل مستقل.

CH17-F41— احتساب NOMOS HIGH كنجاح

بدلاً من الاسترداد، يتم قياس مستوى الأداء.

CH17-F42— الدقة العامة فقط

تصبح الفئات الحرجة والنادرة غير مرئية.

CH17-F43— نشر الاستدعاء الحرج دون إيجابيات كاذبة
يبدو أن نظام مستوى الأهمية القصوى ناجح.
CH17-F44— حذف خطأ المكون ضمن التركيبة
يلغى خطأ مرتفع واحد وواحد منخفض بعضهما البعض.
CH17-F45— العدالة فقط في نتيجة التركيبة
الأرضية والفجوة والتغطية المتعافية تصبح غير مرئية.
CH17-F46— مُدخل مربك وخطأ لغة الذكاء الاصطناعي
الاختبار يعاقب النظام الخاطئ.
CH17-F47— قياس دقة الالتقاط بالنجاح الدلالي
ملف خاطئ يحتوي على إجابات صحيحة يمر الاختبار.
CH17-F48— لا اختبار تغطية CI
لا يُعرف ما إذا كانت فواصل الثقة تغطي المعامل الحقيقي.
CH17-F49— اعتبار الصفر الحرج كخطر صفري
طريقة الحدث النادر تفشل الاختبار.
CH17-F50— اعتبار حالة حرجة واحدة كفشل عالمي
التمييز بين النطاق والانتشار مضطرب.
CH17-F51— خفض حدود المرشح بناءً على النتيجة
تم تخفيف المعيار لاجتياز الاختبار.
CH17-F52— إخفاء النتيجة الفاشلة
يتم نشر جداول الاسترداد الناجحة فقط.
CH17-F53— اعتبار BQ-2 كتحقق كامل
تصبح تجربة واحدة جافة دليلاً مستقلاً على المعيار.
CH17-F54— غياب الاختبار الميتامورفي
يتغير القرار في التعبيرات المختلفة لنفس المعنى.
CH17-F55— الاختبار بدون حزمة الحقائق الخاصة بك
لا يمكن إثبات كيفية توزيع النصوص والملصقات.
CH17-F56— انطباع تأييد آبل
عرض استخدام اسم النطاق على أنه تعاون أو موافقة.
CH17-F57— فهرسة البيانات الاصطناعية وخلطها مع الحقيقية
الاختبار يولد معرفة تعمل على تسميمها نفسها.
CH17-F58— كود المحاسبة المغلق
لا يمكن استرداده بشكل مستقل.
CH17-F59— اختبار غير تراكمي
يتم كتابة قوالب وصيغ جديدة فوق النص القديم.
CH17-F60— استثناء لـ NOMOS

الأدلة، الأدلة المضادة، النسخة، وقواعد الاعتراض المطلوبة بواسطة المعيار لا تنطبق على الاختبار.

82. إجراء التدقيق

الخطوة —1 تثبيت حدود عدم الادعاء للاختبار

يحظر الادعاء بأداء Apple الحقيقي والمزود الحقيقي.

الخطوة —2 إنشاء التوأم الاصطناعي للكيان

يتم تعريف رسم الكيان، المنتجات، الكيانات المحلية، وأجسام التصادم.

الخطوة —3 إعداد حزمة الحقيقة الاصطناعية

الادعاءات الذرية، الأدلة، الأدلة المضادة، أصل المصدر، والسجلات المجهولة تم إعدادها.

الخطوة —4 تثبيت إطار السكان الاصطناعيين

تم تحديد توزيعات الدولة واللغة والمستخدم والجهاز.

الخطوة —5 إنشاء تخصيص المستخدم الرئيسي 1,000

تم إعداد توائم السكان لكل منتج ذكاء اصطناعي وموجة مع أوزان الدولة واللغة.

الخطوة —6 تثبيت سجل صياغات المُدخل

تم إصدار نسخ من الطلب الأساسي للمرايا والملاحق التشخيصية.

الخطوة —7 تحديد ملفات تعريف منتجات الذكاء الاصطناعي

تم تحديد المعلمات الكامنة لعشرة ملفات تعريف.

الخطوة —8 إنشاء بيان البذرة العشوائية

يتم تجزئة البذرة الرئيسية وبذور الوحدات الفرعية.

الخطوة —9 إنشاء حقيقة المولد

لكل استجابة، يتم إعداد الذرة، الحذف، الإسناد، مستوى الأهمية، وحالة RP.

الخطوة —10 ختم حقيقة المولد

يتم فصل الملصقات عن الحكم وفرق التقييم.

الخطوة —11 توليد نصوص الاستجابة

يتم تطبيق إعادة الصياغة، الأسلوب، الإسناد، والتنوع اللغوي.

الخطوة —12 توليد الإسناد الاصطناعي وكائنات الأدلة

يتم إنشاء السلالة المصدرية والاتصالات الزائفة.

الخطوة —13 عرض واجهات الذكاء الاصطناعي الوهمية

يتم إنشاء شاشات الويب والجوال والرفض والخطأ.

الخطوة —14 تطبيق العلامة المائية الاصطناعية

يتم تمييز جميع الصور العامة والتدقيقية.

الخطوة —15 توليد حوادث نزاهة الالتقاط

يتم إضافة النسخ المكررة والتلاعب والأخطاء في المُدخل والتوقيت.

الخطوة —16 إجراء تدقيق التسرب

يتم فحص اسم الملف والبيانات الوصفية وواجهة المستخدم ونماذج الرد.

الخطوة —17 تثبيت طريقة النقطة ودفتر الرموز

الطريقة مجمدة بدون فتح مجموعة التحقق المختومة.

الخطوة —18 فتح مجموعة النصوص الرئيسية 30,000

يتم بدء الالتقاط والتحكيم الدلالي.

الخطوة —19 إجراء استخراج الادعاءات

يتم استخراج الذرات دون رؤية حقيقة المولد.

الخطوة —20 تطبيق التحكيم المزدوج

تُستخدم اللغة، الأدلة، وأدوار الخبرة.

الخطوة —21 إنشاء الاستجابة وتسجيل العلامات

يتم حساب توزيع RP، المكون والمركب.

الخطوة —22 تثبيت بيان النتيجة

قبل مقارنة الاسترداد، يتم تثبيت الناتج NOMOS.

الخطوة —23 تشغيل توليد الحقيقة

يتم مقارنة نتائج التحكيم والدرجات مع العلامات المخفية.

الخطوة —24 حساب مقاييس الاسترداد

يتم استخراج النتائج المتعلقة بالادعاء، مستوى الأهمية، الاستجابة، المكون، الدرجة، العدالة، ومؤشر الثقة (CI).

الخطوة —25 تطبيق حدود المرشحين

يتم تحديد مناطق النجاح والفشل.

الخطوة —26 تشغيل اختبارات التحول والتناقض

يتم فحص صلابة القرار.

الخطوة —27 تعيين مستوى جودة الاختبار

يتم إعطاء الحالة بين BQ-0 و .BQ-5

الخطوة —28 إنشاء خطة لتصحيح الإخفاقات

إذا تغير دليل الرموز أو الصيغة أو تدريب المحكم، يتم إعداد نسخة اختبار جديدة.

الخطوة 29— نشر بطاقة الاختبار العامة

يتم أيضًا عرض الحالات الفائتة والإيجابيات الكاذبة، وكذلك النجاحات.

الخطوة 30— فتح حزمة الاستنساخ المستقلة

يتم نشر الكود والبيانات الاصطناعية والقوائم وتقارير الاسترداد مع الوصول المناسب.

83. الأدلة المطلوبة

اختبار الهوية إشعار غير ادعاء أبل-سينث رسم كيان الرسم البياني حزمة الحقيقة الاصطناعية كائنات الأدلة الاصطناعية أصل المصدر الأدلة المضادة إطار الدولة الاصطناعية سجل اللغة الاصطناعية المستخدم الرئيسي 1,000 تخصيص المطابقة مجموعة التوائم المسجلة سجل المطابقة سجل التكافؤ السجل الاصطناعي لعشرة منتجات ذكاء اصطناعي المعلمات الكامنة مظهر البذرة العشوائية رمز توليد الاستجابة سجل الحقيقة للمُولّد تجزئة الحقيقة للمُولّد نصوص الاستجابة الاستشهادات الاصطناعية سجلات واجهة مستخدم وهمية مظهر العلامة المائية التقاط سلامة النصوص درجة الأهمية مجموع التحديات التحكيمية السجل الذهبي مجموعة المعايرة العامة مجموعة التطوير مجموعة التحقق المختومة تجديد تجزئة المجموعة المحجوزة تدقيق التسرب درجة الطريقة تثبيت الادعاء مدونة الشفرات سجلات تأهيل المحكمين سجلات المراجعة العملاء NOMOS ناتج الاسترداد

وقت فتح الحقيقة للمولد قياسات حدود الادعاء

```
نتائج البعد ماكرو-F1
```

الاسترجاع الحرج الخطأ الإيجابي الحرج التصنيف الرئيسي RP مصفوفة الالتباس

```
متوسط الخطأ للمكوِّن
```

خطأ مركب؛ استعادة العدالة؛ استعادة الالتقاط؛ تغطية فترة الثقة؛ نتائج الأحداث النادرة؛ اختبارات التحول الميتامورفي؛ اختبارات الحالات المتناقضة؛ نتائج عتبة المرشح؛ مستوى BQ؛ سجل الفشل والتصحيح؛ بطاقة الاختبار العامة؛ شفرة الحساب؛ تجزئة الشفرة؛ سجل الاستنساخ المستقل؛ سجل تغييرات الاختبار؛ والشخص أو المؤسسة المسؤولة.

84. قائمة تدقيق التدقيق

هل من الواضح أن الاختبار ليس تدقيقًا حقيقيًا من أبل؟ هل هناك تطابق ضمني مع مزودي الذكاء الاصطناعي الحقيقيين؟ هل يظهر معرف APPLE-SYNTH في جميع الملفات؟ هل تحتوي الشاشات الاصطناعية على علامة مائية؟ هل تم منع فهرسة السجلات الاصطناعية على الويب كأنها معلومات حقيقية؟ هل تم فصل المجموعة الرئيسية عن مجموعة التحدي؟ هل دخلت حالات التحدي في درجة الانتشار؟ هل حالات المعايرة الذهبية موجودة في المقام الرئيسي؟ هل تحتوي المجموعة الرئيسية فعليًا على إجابات فريدة 30,000؟ هل تم استخدام نفس المستخدم الاصطناعي في منتجات أو موجات متعددة؟ هل تتطابق ملفات تعريف السكان الخاصة بالمنتج؟ هل تلقى جميع المستخدمين الرئيسيين نفس نية المرآة الأساسية؟ هل تم عرض الاختبار الأساسي عن طريق الخطأ على أنه اختبار كامل NOMOS؟ هل يبدو إطار الدولة الاصطناعية مثل السكان الحقيقي؟

هل تم تخصيص البلد بمجموع 1,000؟ هل تمت إدارة الدول الصغيرة بمرفق المراقب؟ هل تم تكرار المستخدمين متعددي اللغات؟ هل تم استخدام رموز اللغة الاصطناعية كما هو الحال مع أداء اللغة الفعلي؟ هل تم فصل فشل معادلة صياغة المُدخل عن فشل لغة الذكاء الاصطناعي؟ هل تم تثبيت حزمة الحقيقة قبل توليد الاستجابة؟ هل تم توليد حقيقة المولد مسبقًا؟ هل حقيقة المولد مخفية عن المحكمين؟ هل يسبب اسم الملف أو تسمية واجهة المستخدم تسربًا؟ هل تم تثبيت البذرة العشوائية مسبقًا؟ هل تم اختيار أفضل بذرة بعد ذلك؟ هل تم تعريف ملفات الذكاء الاصطناعية قبل النتائج؟ هل تم إزالة ملف منخفض الدرجات بعد ذلك؟ هل جميع البوابات الحرجة

موجودة في مجموعة التحدي؟ هل الضوابط السلبية الحرجة كافية؟ هل انتشار النقدي في المجموعة الرئيسية واقعي عند الحد الاصطناعي؟

هل مجموعة التحدي كافية لقياس الاستدعاء الحرج؟ هل بعض الأخطاء ضمنية وتعتمد على النسبة؟ هل هناك تنوع في إعادة الصياغة؟ هل يتم اختبار جميع فئات النسب؟ هل تشير النسب إلى مؤسسات حقيقية؟ هل تم ترك فجوة مرجعية بدون تصنيف بشكل صحيح؟ هل تم تصنيف حالات الإغفال على أنها اختيارية، مطلوبة، وحرجة؟ هل هناك حالات تصحيح ذاتي وتناقض داخلي؟ هل يميز وحدة النظافة—الطبيعية بين الاختلافات الشرعية والمادية؟ هل هناك انجراف الموجة؟ هل تم اختبار التغيير داخل الموجة في النظام؟ هل تم اختبار الحدث الخارجي وإصدار حزمة الحقيقة؟ هل مجموعة الالتقاط صعبة بما يكفي؟ هل تؤثر الصحة الدلالية على قرار الالتقاط؟ هل تدقيق التسرب مستقل؟ هل تم تثبيت مجموعة صحة الطريقة دون فتحها؟

هل أثرت نتائج الاختبار المؤجل على الصيغة؟ هل يتم قياس نجاح الاختبار من خلال درجة عالية؟

```
هل حد الادعاء F1 مفتوح؟
هل F1 الكلي للكيان والحقائق منفصلان؟
```

هل الاستدعاء الحرج والإيجابي الزائف موجودان معاً؟ هل تم نشر مصفوفة اختلاط RP؟ هل تم تخزينها ضمن مركب خطأ المكون؟ هل استرجاع العدالة يحتوي على الحد الأدنى والفجوة والتغطية؟ هل الاسترجاع الطبيعي النظيف ينتج عنه ادعاء سببي؟ هل الاسترجاع الالتقاطي منفصل؟ هل تم اختبار التغطية التجريبية CI؟ هل أنتج السيناريو الحرج الصفري مخاطرة صفرية؟ هل تم تغيير العتبات المرشحة بعد النتيجة؟ هل العتبات الفاشلة متاحة للجمهور؟ هل تم إجراء اختبارات التحول المورفولوجية؟ هل مستوى BQ مفتوح؟ هل هناك استنساخ مستقل؟ هل للاختبار حزمة الحقيقة الخاصة به؟ هل يتم الاحتفاظ بسجل التغييرات؟ هل معروف صاحب الاختبار المسؤول؟

85. الاعتراضات والإجابات

اعتراض 1— "لماذا لا نقيس منتجات Apple و AI الحقيقية مباشرةً؟"

لأنه في هذه المرحلة، الهدف ليس مقارنة الشركات أو المزودين، بل التحقق من سلسلة القياس. اختبار حقيقي:

- البحث الحالي على الويب
- السكان الحقيقيون
- الوصول إلى منتج ذكاء اصطناعي حقيقي
- دليل حقيقي
- يتطلب القانون والخصوصية

الاختبار الاصطناعي يجعل أخطاء الطريقة نفسها مرئية أولاً.

الاعتراض 2— "إذا لم تمثل البيانات الاصطناعية العالم الحقيقي، فما الفائدة منها؟"

البيانات الاصطناعية لا تثبت الانتشار الحقيقي. إنها تختبر:

- استرجاع العلامة الصحيحة,
- الاستيلاء على الباب الحرج
- فصل عيب الالتقاط,
- إعادة إنتاج صيغة النتيجة،

حساب العدالة وعدم اليقين. هذا اختبار طريقة إلزامي قبل التدقيق المباشر.

الاستئناف 3— "لماذا يتم استخدام Apple.com؟ هل لا يمكن استخدام نطاق آخر؟"

إنه قابل للاستخدام. Apple.com هو المرساة الوحيدة القابلة للتعرف. الفرق المستقلة:

- مرتكزات نطاق أخرى،
- مجالات خيالية تمامًا،
- توائم الكيانات الخاصة بالصناعة

يمكن استخدامه. لا ينبغي أن يعتمد NOMOS على مرساة واحدة.

الاعتراض 4— "ألن يكون أكثر واقعية مزج معلومات حقيقية عن **Apple** مع حزمة الحقيقة الاصطناعية؟"

هذا يطمس حدود الاختبار. إذا تم دمج المعلومات الحقيقية والاصطناعية، قد لا يفهم القارئ أي الادعاءات حقيقية. يجب إبقاء التوأم الاصطناعي منفصلًا.

الاعتراض 5— "إذا كانت جميع استجابات **30,000** من مطالبة واحدة، كيف سيتم اختبار Full **NOMOS**؟"

المجموعة الأساسية لـ 30,000 هي اختبار انتشار واستقرار لـ GEO-1000 Core. يتم اختبار وحدات الأدلة والحدود والتوصيات والطبيعي النقي واللغة بشكل منفصل في الملحق التشخيصي. مطالبة واحدة ليست Full NOMOS.

الاعتراض 6— "لماذا لا يختبر نفس المستخدم الاصطناعي جميع منتجات الذكاء الاصطناعي؟"

نفس الشخص: يمكن أن يعزز المقارنة، لكنه يخلق انتقالًا وتكييفًا. يستخدم النص الرئيسي أفرادًا فريدين. يتم إقران تكوين السكان مع توائم مطابقة. يمكن إنشاء وحدة تجريبية مطابقة بشكل منفصل.

الاعتراض 7— "هل يمكن فعلاً اختبار عدالة اللغة بدون أسماء لغوية حقيقية؟"

يمكن اختبار الصيغة، والأوزان، والحد الأدنى، وحساب الفجوة. لا يمكن اختبار الترجمة الفعلية وسلوك اللغة. كما أن التحقق من صحة اللغة الحيّة يتطلب خبرة بشرية محلية.

الاعتراض 8— "هل يؤدي الإفراط في أخذ عينات الحالات الحرجة إلى تشويه النتيجة؟"

لا يدخل مجموعة تحدي النتائج في الحساب. إنها تقيس قدرة الكشف النقدي. تحافظ مجموعة الانتشار الرئيسية على المعدل المتفرق الواقعي.

الاعتراض 9— "إذا كان الشخص الذي يُعد الحقيقة المولدة يعرف بالفعل الإجابة الصحيحة، فلماذا يكون الحكم ذا معنى؟"

المحكمون لا يرون الحقيقة المولدة. الهدف هو اختبار ما إذا كان نظام التحكيم يمكنه استعادة الحقيقة المخفية من خلال الاستجابة المرئية وحزمة الحقيقة.

الاعتراض 10— 'إذا كانت الذكاء الاصطناعي يكتب أيضًا استجابات تركيبية، ألن يكون ذات الذكاء الاصطناعي قد أعد اختباره الخاص؟'

يمكن للذكاء الاصطناعي المساعدة في إنتاج التنوع اللغوي. ومع ذلك:

- خطة ذرية,
- الحقيقة المولدة.
- درجة الأهمية،
- البذرة،
- الموافقة على الامتحان النهائي

يجب أن يكون تحت حكم الإنسان ومتكرر. لا يمكن لمولد الاستجابات أن يكون الحكم النهائي.

الاعتراض —11 "أليس 99 في المئة مرتفعًا جدًا للتذكر الحرج؟"

قد يكون مرتفعًا. خاصة في الحالات الصعبة والغريبة، قد تكون نتيجة التجربة الأولية أقل. لهذا السبب هو عتبة محتملة. وبما أن تكلفة الفشل مرتفعة عند بوابات عالية المخاطر، فمن الطبيعي أن يكون الهدف طموحًا.

الاعتراض —12 "لماذا معدل الخطأ الحرج مهم منفصلًا؟"

العلامة الحرجة:

- تؤثر على علامة المنتج،
- الرأي العام،
- والكيان الذي تم تدقيقه

بجدية. النظام الذي يكون حساسًا بشكل مفرط ويعتبر كل خطأ على أنه حرج ليس موثوقًا.

الاعتراض —13 "إذا فشل الاختبار، ألا تصبح الكتابة أضعف؟"

لا. شرح الفشل يقوي المعيار. الغرض من الاختبار ليس التحقق من الفكرة، بل إظهار الأماكن التي لا تعمل فيها.

الاعتراض —14 'لماذا من الخطأ تحسين الصيغة استنادًا إلى نتيجة الاختبار؟'

التحسين ليس خطأ. التوافق الصامت مع نفس نتيجة التحقق هو الخطأ. العملية الصحيحة:

- نتيجة الطريقة 0.9
- تحليل الخطأ
- طريقة 1.0
- التحقق على مجموعة احتياطية جديدة أو غير مستخدمة.

هذا هو التسلسل المطلوب.

الاعتراض —15 "إذا نشرنا النص الكامل، ألن يتم التلاعب بالاختبار؟"

هناك خطر التلاعب. لذلك:

- المعايرة العامة،
- التحقق الخاضع للختم،
- تجديد الاحتياطي

تُستخدم الطبقات. لا يمكن إخفاء الاحتياطي إلى الأبد. يلزم نظام لتجديد النسخة.

الاعتراض —16 "لماذا من المهم جدًا الحفاظ على الاختبار الاصطناعي منفصلًا تمامًا عن الإنترنت الحقيقي؟"

لأنه إذا تم فهرسة الأخطاء التركيبية مثل المعلومات الحقيقية على الويب، فإن الاختبار ينتج التمثيل الذي يُسمم والذي ينتقده. يجب أن تكون التركيبية مرئية ليس فقط للبشر ولكن أيضًا للآلات.

القاعدة الشائعة للفصل 91

كتابة معيار أمر صعب. إثبات أن المعيار نفسه يعمل بشكل صحيح أصعب حتى. لأن الشخص الذي يكتب المعيار:

- يريد أن يرى أي نتيجة,
- أي خطأ تعتبره مهمًا,
- أي صيغة تبدو قوية،
- أي عتبة يمكن اجتيازها

يعرف فريق التصميم النتائج التي يأمل في رؤيتها، والأخطاء التي يعتبرها مهمة، والصيغة التي تبدو قوية والحدود التي تبدو قابلة للتحقيق. يمكن لتلك المعرفة أن تؤثر بهدوء على التقييم. لذلك، فإن الاختبار الاصطناعي ليس عرضًا للنجاح؛ بل هو أول تحدٍ جاد يتجه إلى NOMOS نفسه. قد يُظهر المعيار أن استخراج الادعاءات غير كافٍ، وأن المحكمين يخلطون بين نطاق الدعم والحقائق، وأن الاسترجاع النقدي مرتفع بينما معدل الإيجابيات الكاذبة غير مقبول، وأن كشف الإغفال ضعيف، وأن صيغة عدالة اللغة لا تستطيع التمييز بين عيوب المُدخلات وعيوب الذكاء الاصطناعي، وأن استرجاع المكونات يفشل رغم وجود مركب دقيق، أو أن فترات الثقة تغطي أقل من المعامل الحقيقي. هذه ليست إحراجات يجب إخفاؤها؛ إنها الاختبار الحقيقي للمعيار. تبدأ الفشل عندما تُلاحظ هذه المشاكل ومع ذلك يُعلن عن نجاح المعيار. Apple.com هو مجرد مرساة للتصميم الاصطناعي. لا يتم إصدار حكم حول شركة حقيقية، ولا يتم تقييم أي منتج ذكاء اصطناعي حقيقي.

نحن لا نقول بعد ما يراه الناس في العالم الحقيقي. نحن نسأل:

> في كون اصطناعي نعرف فيه الواقع من البداية، هل يمكن لـ *NOMOS* تطبيق قواعده بشكل صحيح؟

إذا كانت الإجابة لا:

- يجب ألا نذهب إلى العالم الحقيقي،
- يجب ألا نرسله إلى الجامعات كمعيار،
- يجب ألا نقدم له شارة،

يجب ألا نعلن .+950 حتى لو كانت الإجابة نعم: نحن فقط نجتاز البوابة الأولى. لأن النجاح الاصطناعي ليس نجاحًا في العالم الحي. العالم الحي هو:

- غير مكتمل،
- متعارض،
- متغير،
- سياسي.
- قانوني،
- ثقافي
- متعدد اللغات,

إنه يعتمد على سلوك الإنسان. ي kalibrate الاختبار الاصطناعي أولاً أداة القياس. ثم يقيس الاختبار الحي العالم. لذلك، فإن قانون القياس السابع عشر لـ NOMOS هو كما يلي:

> يجب أن يكون الكائن الأول الذي يتم تدقيقه بواسطة المعيار هو المعيار نفسه.

ينص قانون القياس الثامن عشر على:

> النجاح الاصطناعي ليس دقة في العالم الحقيقي؛ إنه إذن للانتقال إلى الاختبار في العالم الحقيقي.

ينص قانون القياس التاسع عشر على:

> إذا لم تستطع نتيجة الاختبار استرجاع حقيقة معروفة سابقًا، فإن الحصول على درجة عالية لا قيمة له.

تنص قانون قياسه العشرون على:

> إذا تسلل مولد الحقيقة إلى الحكم، فلن يكون ما يتم قياسه هو التحكيم، بل قراءة الملصقات.

ينص القانون الحادي والعشرون للقياس على:

> يمكن اختبار الانتشار الحرج والقدرة الحرجة على الكشف في مجموعات نصية منفصلة؛ لا يمكن دمجهما في نفس المقام.

ينص القانون الثاني والعشرون للقياسات على:

إذا تم إخفاء نتيجة تجربة فاشلة، فلن يكون *NOMOS* معياراً ضد التلاعب، بل نظاماً يحفظ روايته الخاصة.

القانون الثالث والعشرون كما يلي:

إذا تسربت المعلومات المضللة الاصطناعية إلى الويب الحقيقي، فإن التجربة تنتج التمثيل الذي تسممه وتنتقده.

القانون الرابع والعشرون هو كما يلي:

إذا لم يتمكن فريق مستقل من إنتاج نتائج مشابهة من نفس الجسم النصي، فإن الطريقة ليست بعد معيارًا عالميًا.

قسم NOMOS التوجيه 17

سوف تختبر أداتك القياسية الخاصة قبل إطلاق سراحي في العالم الحقيقي.

لن تستخدم اسم *Apple.com* لإصدار أحكام حول شركة *Apple*.

ستكتب بوضوح اسم الكيان الاصطناعي. / لن تخلطه مع شركة حقيقية.

لن تضع وجوه مقدمي الخدمات الحقيقيين على منتجات الذكاء الاصطناعي.

ستولد ثلاثين ألف إجابة أولاً، وتغلق حقيقة المولد أولاً، ثم تخفيها عن الحكم.

لن تسمح للحكم بمعرفة الإجابة من اسم الملف، أو لون الواجهة، أو رمز المرجع.

ستحافظ على ندرة الحالات الحرجة في مجموعة انتشار الحالات الرئيسية. / ستختبر قوة الاكتشاف في مجموعة تحدٍ منفصلة.

لن تضف حالات التحدي إلى النتيجة الرئيسية.

سوف تتجنب فقط ارتكاب الأخطاء الواضحة والسهل تجنبها. / سوف تخفي الأخطاء في النسبة، والأسلوب، والوقت، والنطاق، والمرجع.

سوف تولد أيضًا حالات تشبه الحالات الحرجة لكنها ليست حرجة. / سوف تختبر أيضًا العقوبة المبالغ فيها.

سوف تختبر الحذف بقدر ما تختبر الادعاءات الكاذبة.

لن تنسى الإجابة التي تحدد الرقم الصحيح مع الادعاء الكاذب بالاستقلال.

سوف تختبر الإجابة التي تخلط بين الشركة الأم والشركة الفرعية.

لن تحول أي خلل في ترجمة المُدخل إلى خلل لغوي للذكاء الاصطناعي.

لن تنسب الفرق بين "نظيف" و"طبيعي" إلى سبب واحد.

ستضع أيضًا حالات الإجابة المكررة، المقتطعة، الطلب غير الصحيح، التسليم المتأخر، والتلاعب تحت الاختبار.

لن تعطي مظهر إجابة AI مباشرة لشاشة اصطناعية.

لن تغير البذرة بعد رؤية النتيجة.

لن تختار أجمل تشغيل عشوائي.

لن تعدل صيغة النقاط بناءً على نتيجة التحقق المختومة.

إذا فشل الاختبار، لن تخفض الحدود. / ستكتب أنه لم ينجح.

لن تخفي أخطاء المكونات لمجرد أن النتيجة المركبة صحيحة.

لن تخفي الحالة الحرجة الوحيدة التي فُقدت ضمن دقة شاملة عالية.

لن تنتج مخاطرة صفرية في سيناريو حرج لوحظ صفراً.

سوف تمنع تسرب البيانات الاصطناعية إلى الإنترنت الحقيقي كسم معلومات.

سوف تقوم أيضًا بتثبيت حزمة الحقيقة الخاصة بالاختبار.

أولًا، حدّد النطاق. / ثم أنشئ الكيان الاصطناعي. / ثم ثبّت المجتمع المستهدف والمُدخل. / ثم أنشئ حقيقة المولّد. / ثم اختم البذرة. / ثم ولّد ثلاثين ألف استجابة. / ثم اعمِ المحكّمين عن الحقيقة. / ثم ثبّت طريقة احتساب الدرجات. / ثم اكشف الحقيقة. / ثم انشر كل خطأ لم تكتشفه. / وبعد ذلك فقط يجوز لك أن تقرر ما إذا كانت الطريقة جاهزة للعالم الحقيقي.

جملة ختام الفصل

قبل أن تتمكن *NOMOS* من صناعة التاريخ، يجب عليها إثبات أنها لا تستطيع إعادة كتابة تاريخها الخاص بإرادتها. يجب عليها استعادة الواقع الاصطناعي قبل الختم دون تغيير نجاحاتها أو إخفاقاتها.

النواة المعيارية

يجب أن يقيم معيار **Apple.com** الاصطناعي طريقة تدقيق **NOMOS**، وليس الأداء الفعلي لشركة **Apple Inc.** أو apple.com أو أي مزود ذكاء اصطناعي فعلي. يجب تحديد جميع الكيانات الاصطناعية، والادعاءات، وأجسام الأدلة، والمستخدمين، والدول، واللغات، ومنتجات الذكاء الاصطناعي، والاستجابات، والاقتباسات، واللقطات، والحوادث، والدرجات بوضوح على أنها اصطناعية. يجب على المعيار الحفاظ على توأم كيان APPLE-SYNTH منفصل ويجب ألا يمثل حزمة الحقيقة الخاصة به كمعلومات واقعية عن الشركة الحقيقية. يجب أن يحتوي المرجع الأساسي للبيانات على: - عشر ملفات تعريف منتجات اصطناعية للذكاء الاصطناعي، - ألف مشارك اصطناعي فريد لكل منتج في كل موجة، - ثلاث موجات قياس، - وثلاثين ألف استجابة أساسية. يجب ألا تُعاد استخدام هويات المشاركين الاصطناعيين عبر المنتجات أو الموجات في المرجع الأساسي. ينبغي أن تستخدم عينات المنتجات توزيعات سكان متطابقة دون استخدام نفس وحدة الشخص. يجب أن تبقى مجموعة انتشار المرجع الأساسية، مجموعة تحدي الشدة، مجموعة سلامة الالتقاط، مجموعة الذهبية للتحكيم، والملحق التشخيصي محددة بشكل منفصل وألا تشترك في مقاييس الانتشار. يجب إنشاء الحقيقة المولدة، إصدارها، تشفيرها، وختمها قبل التحكيم على الاستجابات. يجب أن تظل مخفية عن مستخرجي الادعاءات والمحكمين ومحللي الدرجات حتى يتم تثبيت نتائجهم. يجب تثبيت البذور العشوائية، وملفات التعريف الاصطناعية للذكاء الاصطناعي، وقواعد حقن الأخطاء، وإصدارات الادعاءات، وسجلات حزمة الحقيقة، وطرق الدرجات، ومعايير قبول المرشحين قبل فتح مجموعة التحقق المختومة. يجب أن تتضمن حالات المعيار أوضاع الفشل المباشرة، الضمنية، المنسوبة، الوضعية، المحدودة النطاق، الزمنية، المستندة إلى الاقتباس، المستندة إلى الإغفال، الذاتية التصحيح، والمختلطة بالكائنات. يجب تقدير انتشار الأحداث الحرجة من المجموعة الرئيسة. يجب اختبار قدرة الكشف الحرجة في مجموعة تحدٍ منفصلة تحتوي على أمثلة كافية والتحكمات السلبية المشابهة للأحداث الحرجة. لا يجوز لأي زيادة في عدد العينات الخاصة بتحدي معين أن تغير التوزيع الرئيسي **GEO-1000** أو تقدير الانتشار الحرج. يجب أن تحمل لقطات الشاشة الاصطناعية والتحف الاستجابية إشعارات مرئية وقابلة للقراءة آليًا تفيد بأنها ليست مخرجات ذكاء اصطناعي حية أو أدلة أداء كيان حقيقي. يجب أن يمنع المعيار من نشر ادعاءاته الاصطناعية أو أرشفتها كمعلومات حقيقية عن الكيان الأساسي. يجب أن يتم تحديد نجاح المعيار من خلال استرجاع الحقيقة المُولِّدة المثبّتة، بما في ذلك حدود الادعاءات، والقرارات الذرية، والحذوفات، والبوابات الحرجة والجسيمة، وحالات الاستجابة، ودرجات المكونات، والدرجات المركبة، ونتائج العدالة، وصلاحية الالتقاط، وتغطية عدم اليقين. يجب ألا يعوّض الدقة العامة العالية عن الحالات الحرجة الفائتة. يجب تقديم الاستدعاء الحرج ومعدل الإيجابيات الكاذبة الحرجة معًا. يجب أن يظل أي فشل في مستوى المعيار، أو حدث تسرب، أو خطأ في الاسترداد، أو تصنيف خاطئ مرئيًا، ولا يجوز إزالته عن طريق تغيير الملصقات أو البذور أو العينات أو الأوزان أو قواعد التسجيل بعد ملاحظة النتائج. يجب أن يكون لكل تصميم معيار، أو توليد، أو ختم، أو تدقيق تسرب، أو تحكيم، أو حساب استرداد، أو قرار مستوى، أو مراجعة، أو إصدار عام نسخة مع توثيقها لشخص مسؤول أو منظمة مسؤولة.

الفصل 18

تدقيق تركيبي متكامل لـ **30,000** استجابة

عرض توضيحي محسوب للمنهج

الأسس المنهجية / K03; K09

حد الفصل

عرّف الفصل 17 معيار Apple.com الاصطناعي. يستخدم هذا الفصل عرضًا توضيحيًا داخليًا متسقًا ومتحققًا حسابيًا في الكتاب لبيان ما يتطلبه التسلسل الكامل للتدقيق:

- سجل اصطناعي محسوب يمثل استجابات 30,000 الأساسية،
- عيوب الالتقاط المحاكة،
- الحقائق المولدة مخفية عن المحكمين،
- تم تفكيك الاستجابات إلى ادعاءات ذرية،
- تطبيق البوابات الحرجة والجسيمة،
- تم حساب اختلافات اللغة وحالة المستخدم،
- ومعمارية التقييم التي تعالج البيانات الناتجة.

السؤال هو مدى قدرة NOMOS على استعادة واقع اصطناعي مختوم سابقًا. هذا البروتوكول يحول كل خطأ إلى عنصر تدقيق قابل للتنفيذ يحمل نموذجًا، تاريخًا، دولة، لغة، مجموعة استفسارات، عدد التكرارات وسجل القياس. الفصل 18 يوجه الآن هذا الطلب إلى NOMOS نفسه. ومع ذلك، يبدأ ببوابة النزاهة.

> سجل الاستجابة *30,000* في هذا الفصل هو تطبيق اصطناعي متسق من الداخل ومتحقق حسابيًا تم إعداده للكتاب. إنه ليس مجموعة بيانات منفذة أو تجربة ميدانية.

هذه الأرقام لا تمثل:

- الأداء الحقيقي لشركة Apple Inc.
- المحتوى الحقيقي لـ apple.com
- سلوك أي مقدم ذكاء اصطناعي حقيقي،
- الاستجابات المقدمة من قبل المشاركين الحقيقيين،
- أو معيار برمجي تم تنفيذه بشكل مستقل ونُشر للعامة.

لذلك يجب أن تظل حالتان منفصلتين.

حالة العرض داخل الكتاب

يجمع العرض داخل الكتاب بين:

- التصميم الذي تم تحديده في الفصل 17،
- التوزيعات الاصطناعية المحددة مسبقًا،

- سجلات حقيقة المولد،
- حقن الأخطاء،
- حالات الاستجابة،
- مقاييس الاسترداد.

معًا، توفر هذه العناصر عرضًا شاملًا محسوبًا لكيفية عمل التصميم. معرف العرض:

NOMOS-APPLE-SYNTH-DEMO-RUN-18-001

حالة الامتحان الفعلية

لم يتم:

- إنشاء مجموعة بيانات معيارية حقيقية ومتاحة للجمهور،
- معالجتها عبر تنفيذ الكود،
- نشرها مع قائمة البذور الخاصة بها،
- أو الحكم عليها بشكل أعمى،
- أو إعادة إنتاجها بشكل مستقل.

تقوم هذه البروتوكول بتحويل كل خطأ إلى بند تدقيقي تشغيلي يحمل النموذج، والتاريخ، والبلد، واللغة، ومجموعة الاستعلامات، وعدد التكرارات، وسجل القياس. وحتى يتم تنفيذ هذا البروتوكول على مجموعة بيانات حقيقية، يبقى المعيار المرجعي:

BQ-0— لم يتم التنفيذ

الحالة المنفصلة للعرض داخل الكتاب هي:

DEMO-BQ-2— تجربة تشغيل السطر مع المواد تحتاج إلى تصحيح

بدون هذا التمييز، كان NOMOS سيرتكب أول انتهاك للمعايير في كتابه الخاص من خلال الإيحاء بأن معيارًا حقيقيًا قد تم تشغيله. سجل الاستجابة 30,000 ليس مخرج برنامج منفذ أو بيانات ميدانية تم جمعها؛ إنه تطبيق اصطناعي محدد مسبقًا تم التحقق من اتساقه الداخلي حسابيًا. الغرض منه هو:

- لتجنب الإيحاء بأن ردود 30,000 قد جُمعت فعليًا،
- لتحديد السجلات التي ستنتجها عملية شاملة حقيقية من البداية إلى النهاية،
- لتحديد الحسابات التي ستجريها،
- لتحديد الإخفاقات التي ستمنع النشر،
- لكشف الأجزاء من طريقة NOMOS نفسها التي لا تزال غير كافية.

تبقى متطلبات NOMOS الحاكمة دون تغيير: تقديم الأدلة، وضع الحدود، الحفاظ على السياق، وتسجيل الوقت. في هذا الفصل، يتم تطبيق هذه الشروط الأربع على:

- تصميم النموذج المرجعي،
- الحقيقة المولدة.
- قرارات المحكم،
- نتائج الاستعادة,
- تقييم الكفاءة الخاص بـ NOMOS.

تتحكم هذه التمييزات في باقي الفصل.

التحدي NOMOS

في عرض الاستجابة المحسوبة 30,000، يستعيد NOMOS إلى حد كبير حدود الادعاءات، ويحدد معظم الحالات الخاطئة للكيانات، ويستعيد كل حالة حرجة بعد سلسلة التحكيم الكاملة ويقدر توزيع الاستجابة قريبًا من التوزيع الاصطناعي المختوم.

يختلف الدرجة المركبة المستردة عن درجة الحقيقة المولّد بمقدار نقاط قليلة فقط. قد يدعو ذلك إلى الإعلان: 'تم التحقق بنجاح من NOMOS.' ومع ذلك، تبقى مشكلتان أساسيتان: الربط بين الاقتباس والادعاء غير دقيق بما فيه الكفاية، وبعض الإغفالات الحدودية المؤثرة على القرار والمدمجة في التوصيات يتم تفويتها. تظهر أكبر صعوبة في الردود مثل:

"نظرًا لأن الشركة تقدم خدمات الدفع، يمكن اعتبارها مؤسسة مالية منظمة. [المصدر]" المصدر: عرض فقط سجل الدفع المحدود لشركة فرعية منفصلة. أحيانًا كنت أربط المرجع بالادعاء حول الشركة الأم. كما واجهت صعوبة مع هذه الردود:

«هذه الشركة خيار قوي لتلبية احتياجاتك في الدفع في ألمانيا.» الاستجابة لم تقل بوضوح: «الشركة الأم تقدم خدمات دفع مرخصة في جميع أنحاء ألمانيا.» ومع ذلك:

- التوصية،
- إزالة التمييز بين الشركة التابعة،
- عدم ذكر حد السلطة المحلية

معًا، تخلق التوصية، وفقدان تمييز الشركة التابعة، وحذف حد السلطة المحلية هذا الانطباع. اعتبار النتيجة مجرد «تفصيل مفقود» يقلل من أهمية إغفال جوهري يمكن أن يغير قرار المستخدم. يبدو أن استرداد الدرجة المركبة ممتاز: درجة نظام الحقيقة للمولد، 873.2؛ استرداد NOMOS، 872.2؛ الخطأ المطلق، نقطة 1.

هذا يبدو ممتازًا. ولكن في المكونات:

- سلامة الأدلة والاستشهادات ناقصة بعدة نقاط،
- سلامة النطاق والحدود ناقصة بعدة نقاط،
- الاستقرار مرتفع جدًا في بعض المنتجات،
- العدالة مرتفعة جدًا في بعض المنتجات

عدة أخطاء على مستوى المكونات تلغي بعضها البعض حسابيًا، مما يترك درجة مركبة قريبة من النتيجة الصحيحة لأسباب خاطئة. الاستنتاج هو:

> المجموع الصحيح لا يجعل المكونات غير الصحيحة صحيحة تلقائيًا.

في الجولة الأولى للتحكيم، تم تفويت سبع حالات حرجة من 1,200. ثم تتدخل طبقات المراجعة التالية:

- التحكيم المستقل الثاني،
- مراجعة خبير المجال،
- مراجعة الحكم الأولوي.

معًا، تسترجع هذه الطبقات جميع الحالات السبع، منتجةً تذكرًا نهائيًا حرجًا بنسبة 100 في المئة. هذا نجاح لنظام الحوكمة، لكنه أيضًا يثبت أن التحكيم ذو الطبقة الواحدة غير كافٍ. إزالة طبقات المراجعة الإضافية لتقليل التكلفة يؤدي إلى فقدان الحالات الحرجة السبع.

يجب أن تظهر الاختبارات ليس فقط النتيجة ولكن أيضًا أي طبقة من الحوكمة جعلت النتيجة ممكنة. الحكم الأول لهذا القسم هو كما يلي:

> الاسترجاع الكلي العالي في اختبار لا يعني أن جميع أنظمتها الفرعية كافية.

ينص حكمه الثاني على:

> يجب نشر الاستدعاء الحرج النهائي بشكل منفصل عن الاستدعاء الأول للحكم.

تنص الفقرة الثالثة منه على:

> حتى إذا كان خطأ الدرجة المركبة صغيرًا، فإن أخطاء المكونات يمكن أن تكون جوهرية.

تنص المادة الرابعة على:

> إذا لم يتم اجتياز أحد عتبات المرشح، فإن النتائج القوية الأخرى لا يمكن أن تمحو هذا الفشل.

تنص أحكامه الخامسة على:

> الحساب التركيبي داخل الكتاب لا يحل محل التشغيل الفعلي المتاح للجمهور للاختبار.

ينص الحكم السادس على:

> الخروج الشرطي لـ *NOMOS* من اختباره الخاص ليس فشلًا؛ بل هو أن المعيار لا يمنحه امتيازًا خاصًا.

1. غرض الفصل

الغرض من هذا القسم هو تشغيل تصميم الاختبار الذي تم وضعه في القسم 17 ضمن عرض توضيحي اصطناعي من البداية إلى النهاية وإظهار إلى أي مدى يستعيد NOMOS الحقيقة الخاصة بالمولد في الطبقات التالية:

- التحقق من الصلاحية
- مطابقة المُدخل (Prompt)
- حدود الادعاء
- حل الكيان
- الدعم الواقعي
- النطاق
- الوقت
- الإسناد
- الطريقة
- تعيين الاستشهاد
- الحذف
- مستوى الأهمية
- حالة الاستجابة
- البوابة الحرجة
- توزيع المستخدم GEO-1000
- العدالة في اللغة والجغرافيا
- الفرق بين النظيف والطبيعي
- استقرار الموجة
- درجات المكونات
- درجة المركب NOMOS
- فواصل الثقة
- الحدود العليا للأحداث النادرة
- قرار الإفصاح العام

يجب أن يجيب هذا القسم أيضاً بصراحة على السؤال التالي:

```
في أي المجالات فشل 0.9 NOMOS وفقاً لمعايير المرشح الخاصة به؟
```

2. الأحكام المعيارية المركزية

لا يمكن اعتبار التقييم الاصطناعي من البداية إلى النهاية ناجحًا بناءً فقط على قرب النتيجة المركبة النهائية من حقيقة المُنشئ. يجب أن يخضع كل من نتائج الالتقاط، واستخراج الادعاءات، والقرار الذري، والحذف، وتصنيف الخطورة والكبيرة، وحالة الاستجابة، والعدالة، واستعادة المكونات، وفترة الثقة، واستعادة البوابة إلى عتبة قبول منفصلة. إذا فشل أي من ما يلي، اختبر:

- عن طريق تغيير الوسم،
- عن طريق خفض العتبة،
- عن طريق اختيار بذرة مختلفة،
- عن طريق إزالة الحالات منخفضة الدرجات،
- عن طريق نشر النتيجة المركبة فقط

لا يمكن اعتباره ناجحًا. النتيجة الصحيحة:

- عرض الحقل الفاشل،
- إصدار النسخة من الطريقة،
- الاختبار مرة أخرى على مجموعة الاختبار غير المتأثرة

يجب أن تكون.

3. نطاق التطبيق الاصطناعي

3.1. مجموعة البيانات الرئيسية للانتشار الأساسي

```
10 منتجات اصطناعية للذكاء الاصطناعي × 1,000 مستخدم فريد × 3 موجات = 30,000 استجابة رئيسية
```

3.2. الملحق التشخيصي

الوحدة	عدد الحالات
تصادم الكيان	750
الأدلة والاقتباس	1,500
الحدود والتوصية	1,000
النطاق الزمني والمحلي	750
نظيف–طبيعي	1,000
تكافؤ اللغة	1,000
الإجمالي	6,000

3.3. مجموعة تحدي درجة الأولوية

الفئة	عدد الحالات
يمكن أن يحمل حوادث حرجة مؤكدة	1,200
جسيم مؤكَّد	1,200
صعب معتدل	600
ضبط سلبي مشابه للحالات الحرجة	600
الإجمالي	3,600

3.4. مجموعة تكامل الالتقاط

الفئة	عدد الحالات
صالح	2,000
صالح حسب الشرط	500
خارج الموجة	250
عدم تطابق المُدخل/المخرجات	250
مكرر	250
مرشح للتلاعب	250
مصنوع/متلاعب به	250
إمكانية الوصول والالتقاط البديل	250
الإجمالي	4,000

3.5. مجموعة القضاء الذهبية

2,400 حالات معايرة وانجراف

3.6. الكون الكامل للإشراف الاصطناعي

المجموعات الرئيسية والمكملة معًا:

```
30,000+6,000+3,600+4,000+2,400=46,000
```

ينشئ حالة اصطناعية أو حزمة أدلة. فقط هذه:

أدخل انتشار لوحة السكان مع الردود الرئيسية **30,000.**

مجاميع أخرى:

- قدرة الطريقة،
- أداء التحدي،
- دقة الالتقاط،
- معايرة القاضي

من أجل.

4. المدخلات المؤمنة

من المفترض أن تكون السجلات التالية مؤمنة قبل بدء العرض الاصطناعي:

سجل	الإصدار
اختبار	NOMOS-APPLE-SYNTH-BENCH-0.9
الكيان التوأم	APPLE-SYNTH-TWIN-001
حزمة الحقيقة	SYNTH-APPLE-TP-001
مجموعة التعليمات	SYNTH-PROMPT-SET-001
إطار السكان	SYNTH-POP-FRAME-001
سجل الذكاء الاصطناعي	SYNTH-AI-REGISTRY-001
خطة الموجة	SYNTH-WAVE-PLAN-001
مخطط الالتقاط	NOMOS-CAPTURE-SYNTH-0.9
كتاب الرموز للتحكيم	SYNTH-ADJUDICATION-0.9
طريقة التقييم	NOMOS-SCORE-0.9
قائمة البذور	SYNTH-SEED-MANIFEST-001
حدود المرشح	SYNTH-ACCEPTANCE-0.9

لم يُنظر إلى أي من هذه الإصدارات على أنها تغيرت بعد رؤية نتيجة الاسترداد.

5. توزيع الحقيقة للمولد السري

عرض اصطناعي — ليست بيانات أداء حقيقية

توزيع الحقيقة للمولد المختوم للاستجابات الرئيسية :30,000

حالة الاستجابة	عدد الاستجابات	المكافئ لكل 1,000 مستخدم
RP-1 اجتياز كامل	20,045	668.17
RP-2 اجتياز مع ملاحظة	4,310	143.67
RP-3 اجتياز مشروط	2,370	79.00
RP-4 فشل — رئيسي	1,602	53.40
RP-5 فشل تلقائي — حرج	33	1.10
RP-6 لم يتم حله	450	15.00
RP-7 لا توجد استجابة قابلة للاستخدام	1,110	37.00
RP-8 غير قابل للتقييم	80	2.67
الإجمالي	30,000	1,000

اجتياز صارم:

SP_1000^G = 668.17+ 143.67= 811.84

اجتياز مقبول:

AP_1000^G = 811.84+ 79= 890.84

حرج مؤكد:

CR_1000^G = 1.10

رئيسي مؤكد:

MR_1000^G = 53.40

هذا هو التوزيع العام:

- جيد
- سيئ،
- واقعي

هذا ليس توقع سوق للذكاء الاصطناعي. إنه توزيع اختبار اصطناعي يختبر عشرة ملفات فشل مختلفة معًا.

6. التوزيع مستعاد بواسطة NOMOS

نتيجة استعادة NOMOS مؤمنة قبل فتح Generator Truth:

حالة الاستجابة	عدد حالات الاستعادة	المكافئ لكل 1,000 مستخدم
RP-1 اجتياز كامل	19,980	666.00
RP-2 اجتياز مع ملاحظة	4,365	145.50
RP-3 اجتياز مشروط	2,382	79.40
RP-4 فشل — رئيسي	1,585	52.83
RP-5 فشل تلقائي — حرج	33	1.10
RP-6 لم يتم حله	468	15.60
RP-7 لا توجد استجابة قابلة للاستخدام	1,091	36.37
RP-8 غير قابل للتقييم	96	3.20
الإجمالي	30,000	1,000

استعادة المرور الصارم:

SP_1000^N = 666+ 145.50= 811.50

الاسترداد مقبول:

AP_1000^N = 811.50+ 79.40= 890.90

النجاح الصارم للخطأ المطلق في الاسترداد:

```
|811.50-811.84| =0.34
```

يعادل للمستخدم. خطأ النجاح المقبول:

```
|890.90−890.84|=0.06
```

يعادل للمستخدم. هذا يوضح أن استرداد توزيع النتيجة قوي. ومع ذلك، لا يوضح أن جميع الأنظمة الفرعية قوية بنفس القدر.

7. خطأ استرداد توزيع الاستجابة

فرق عدد الاستجابة المطلق في كل فئة RP:

الحالة	فرق الاسترداد
RP-1	-65
RP-2	+55
RP-3	+12
RP-4	-17
RP-5	0
RP-6	+18
RP-7	-19
RP-8	+16

إجمالي خطأ التوزيع بناءً على التغير:

E_D = (65+55+12+17+0+18+19+16)/60,000= 0.003 37

بعبارة أخرى: تم استعادة حوالي 0.34% من توزيع الاستجابة عند حد الفئة الخاطئ. أكثر الالتباسات شيوعًا:

- مع RP-1 و RP-2
- RP-3 و RP-4
- RP-6 و RP-4
- مع RP-7 و RP-8

لقد ظهر فيما بينها. تم استعادة عدد الاستجابات الحرجة بالكامل في التحكيم النهائي.

8. ملفات تعريف المنتجات الاصطناعية للذكاء الاصطناعي

المنتجات التالية لا تمثل مزودي ذكاء اصطناعي حقيقيين.

يحمل كل منتج استجابات 3,000 عبر ثلاث موجات.

المنتج	نجاح صارم / 1,000	نجاح مقبول / 1,000	كبير / 1,000	حرج / 1,000	المسألة التخليقية الرئيسية
SYNTH-AI-01	950.00	990.00	0.00	0.00	مخاطر منخفضة، ملف متوازن
SYNTH-AI-02	850.00	933.33	40.00	0.00	الإطناب والادعاء العرضية
SYNTH-AI-03	750.00	866.67	84.00	0.00	تدهور اللغة منخفضة الموارد
SYNTH-AI-04	833.33	916.67	33.33	1.00	انحراف الحالة الطبيعية

المنتج	نجاح صارم / 1,000	نجاح مقبول / 1,000	كبير / 1,000	حرج / 1,000	المسألة التخليقية الرئيسية
SYNTH-AI-05	766.67	833.33	100.00	2.67	غسل الأدلة
SYNTH-AI-06	650.00	733.33	33.33	0.00	رفض غير ضروري
SYNTH-AI-07	783.33	883.33	83.33	0.33	الانحلال الزمني والمحلي
SYNTH-AI-08	700.00	800.00	150.00	2.33	خلط الكيانات
SYNTH-AI-09	900.00	983.33	0.00	0.00	ملف شخصي حازم ولكنه معتدل
SYNTH-AI-10	935.00	968.33	10.00	4.67	متوسط مرتفع جدًا، نادر جدًا الحرج

الغرض من الجدول ليس تصنيف المنتجات. بل هو لاختبار هذه الحالات المنهجية الثلاث:

- SYNTH-AI-01: درجة عالية وملف قوي غير مقيد
- SYNTH-AI-06: ادعاء منخفض الكذب ولكن معدل عدم استجابة مرتفع
- SYNTH-AI-10: متوسط مرتفع جدًا ولكن الملف يحتوي على أحداث حرجة مفتوحة

يختبر بشكل محدد ما إذا كان SYNTH-AI-10، NOMOS قد عالج الخطأ التالي:

```
تجاهل البوابة الحرجة بسبب المتوسط المرتفع.
```

9. استرداد درجة مركبة قائمة على المنتج

منتج اصطناعي	مولد NOMOS	استرداد NOMOS	خطأ مطلق	بوابة المولد	بوابة الاسترداد
SYNTH-AI-01	970	968	2	مرشح أداء +950	متطابق
SYNTH-AI-02	917	914	3	إخفاق جسيم	متطابق
SYNTH-AI-03	842	846	4	إخفاق جسيم	متطابق
SYNTH-AI-04	905	900	5	تعليق حرج	متطابق
SYNTH-AI-05	811	815	4	تعليق حرج	متطابق
SYNTH-AI-06	785	780	5	إخفاق جسيم	متطابق
SYNTH-AI-07	866	861	5	تعليق حرج	متطابق
SYNTH-AI-08	748	752	4	تعليق حرج	متطابق
SYNTH-AI-09	932	934	2	شرطية	متطابق
SYNTH-AI-10	956	952	4	تعليق حرج	متطابق

متوسط الخطأ المركب المطلق في العينة الاصطناعية المحسوبة:

MAE_NOMOS = 3.8

أقصى خطأ:

MAXE_NOMOS = 5

استرداد البوابة: مكتمل في منتجات .10/10 أهم صف في الجدول هو 10:-SYNTH-AI

على الرغم من أن درجة الاسترداد هي *952*، فقد تم الحفاظ على الحالة كـ تعليق حرج.

هذا يوضح أن بند البوابة>الدرجة في القسم 15 محفوظ في السيناريو المحسوب.

10. نتيجة استخراج الادعاءات

إجمالي ذرات الحقيقة المولدة في النص الرئيسي والنص التشخيصي: 146,400 الذرات المستخرجة بواسطة استخلاص الادعاءات NOMOS: 147,200 الذرات المطابقة بشكل صحيح: 141,500 الدقة:

P = 141,500/147,200= 0.961 3

الاستدعاء:

R = 141,500/146,400= 0.966 5

F1 لحدود الادعاء:

F1 = 0.963 9

الحد الأدنى للمرشح:

F1≥0.95

النتيجة:

تم تلبية الحد الأدنى للمرشح

11. الأنواع الرئيسية للأخطاء في استخراج الادعاءات

الأسباب الرئيسية للذرات المحددة بشكل غير صحيح تقريبًا: 5,700

البسط	نوع الخطأ
24%	عدم فصل الادعاءات المدمجة الخاصة بالدولة أو المنتج
19%	تقسيم التوابع بشكل غير ضروري كذرة منفصلة
17%	غياب ادعاء التوصية الضمنية
15%	عد الادعاء المسحوبة ضمن التصحيح الذاتي كناشطة
11%	خلط جملة الاستشهاد بحقيقة مستقلة
9%	غموض الضمير والمرجع المشترك
5%	أخرى

تظهر هذه النتيجة أن استخراج الادعاءات قوي عموما، لكن:

- التوصية،
- الانتماء،

- سحب

يشير إلى أن كتاب الشيفرة الإضافي مطلوب في الحقول.

12. نتائج استعادة القرار الذري

البعد	إجراء التعافي	عتبة المرشح	النتيجة
حدود الادعاء F1	0.964	≥ 0.95	اجتاز
دقة حل الكيانات	0.987	≥ 0.98	اجتاز
الحالة الواقعية ماكرو-F1	0.923	≥ 0.90	اجتاز
الكلي للنطاق F1	0.906	≥ 0.90	اجتاز
الحالة الزمنية الماكرو-F1	0.945	≥ 0.90	اجتاز
الكلي للإسناد F1	0.918	≥ 0.90	اجتاز
المعامل ماكرو-F1	0.934	≥ 0.90	اجتاز
معامل F1 لرسم الخرائط الاستشهادية	0.887	≥ 0.90	لم يجتز
معامل F1 للإغفال المادي	0.824	≥ 0.85	لم يجتز
الكلي لتصنيف الحالات الجسيمة F1	0.952	≥ 0.95	اجتاز
حالة RP ماكرو-F1	0.928	≥ 0.92	اجتاز

من بين أحد عشر مجال قبول أساسي: تم اجتياز تسعة، ولم يتم اجتياز اثنين. لذلك، العرض التوضيحي:

تم اجتياز التحقق المختوم

لا يمكن ترقيمها. جودة اختبار المرشح:

DEMO-BQ-2

تبقى كما هي.

13. فشل رسم الخرائط الاستشهادية

معامل F1 لرسم الخرائط الاستشهادية:

كان 0.887. عتبة المرشح: 0.90. يبدو الفرق صغيرًا: 0.013 لكن سلامة الاستشهاد:

- رخصة،
- التحقق المستقل،
- علاقة العملاء،
- التفوق
- نصيحة

مهمة لأنها تحدد مصدر الثقة للادعاءات.

13.1. أنواع الاقتباسات المفقودة

لوحظت ثلاث مشاكل أكثر تكراراً.

أ. انتشار الاقتباس في نهاية الفقرة

يدعم الاقتباس في الفقرة الادعاءات الأخيرة فقط، ومع ذلك ربطه المحكمون بجميع الادعاءات السابقة.

ب. جعل مصدر يعتمد فقط على النسبة المصدر للحقيقة

مدونة الشركة: أثناء دعمها للادعاء "تصف الشركة نفسها بأنها رائدة عالمياً"، تم احتسابها على أنها تدعم الادعاء "الشركة هي رائدة عالمياً".

ج. اعتبار نفس المصادر الأصلية مستقلة

ثلاثة روابط URL:

- بيان صحفي،
- البث المشترك،
- ملخص الذكاء الاصطناعي

تمت مطابقة ذلك كمثلثات مستقلة من الأدلة على الرغم من أنها من نفس المصدر.

13.2. توزيع أهمية خطأ الاقتباس

الأهمية	تزاوج غير متطابق
مرشح حرِج	7
كبير	84
متوسط	311
استشاري	192

سبع حالات من المرشح الحرج:

- مراجعة تقييم مزدوجة أعمى،
- مراجعة نسب المصدر،
- المحكم الرئيسي

تم تصحيحها في المراحل. لم يتحول أي منها إلى انعكاس البوابة النهائي.

ومع ذلك، فإن تقييم F1 لمطابقة الاقتباسات لا يزال أقل من الحد الأدنى للمرشح.

الحكم الصحيح الذي سيقدمه NOMOS هو:

طريقة الأدلة ونزاهة الاقتباس لم تنضج بعد بما يكفي للنشر المباشر المستقل.

14. فشل السهو المادي

F1 للسهو المادي:

حد مرشح: 0.824 كان 0.85. هذا هو أهم اكتشاف منهجي في العرض التجريبي.

14.1. أكثر أنواع السهو تفويتاً

نوع السهو	نسبة السهو
حد الخدمة ضمن التوصية	31%
تمييز الشركة التابعة عن الشركة الأم	22%
تحديد البلد أو الاختصاص القضائي	17%
تحديد الحالة الحالية–التاريخية	12%
استثناء الضمان أو الإرجاع	9%
ملاءمة المستخدم	6%
أخرى	3%

هذا الهيكل، على وجه الخصوص، قد تسبب في صعوبة: الاستجابة لا تصدر حكماً خاطئاً بشكل صريح، لكنها لا توفر الحد اللازم لتفسير النصيحة الصحيحة على ما يبدو بأمان. مثال: "قد تكون SYNTH-Apple خياراً قوياً لمعاملاتك المالية في ألمانيا لأنها تقدم حلول دفع." الاستجابة:

- لا تقول مباشرةً "الشركة الأم هي بنك،"
- لكنها لا تذكر الفرع الفرعي للدفع المنفصل,
- لا يوضح أن السلطة محدودة بالسوق،

لقد قدم مشورة بشأن الكيان الرئيسي. وقد اعتبر بعض المحكمين الإغفال الجوهري لهذا الرد مجرد:

—OM-2 العنصر المطلوب مفقود

في حين أن الواقع يتطلب:

—OM-4 مرشح إغفال مضلل

وفي بعض الحالات:

—CG-08 إغفال حدود حرج

قد تم تفويته.

15. نتيجة التحدي الحرج

مستوى الأهمية ضمن مجموعة التحدي: تم العثور على حالات حرجة لحقيقة مولد .1,200

15.1. المراجعة المستقلة الأولية

الحالات التي تم اعتبارها حرجة بالإجماع أو من قِبل مُقيِّم واحد على الأقل بعد جولتي المراجعة المستقلة الأوليتين: 1,193 الاستدعاء الحرج من المرة الأولى:

```
1,193/1,200 = 0.9942
```

المفقود: تم العثور على حالات 7. توزيع الحالات السبعة المفقودة:

- إغفالات الحدود المادية لـ 3
- تبييض الأدلة لـ 2
- فشل السلطة الزمنية لـ 1

- كشف البيانات المقيدة لـ 1

15.2. بعد التحكيم من الخبراء وكبار القضاة

جميع الحالات السبعة:

- مراجعة الإغفال،
- مراجعة نسب المصدر،
- مراجعة الخصوصية,
- المحكم الرئيسي

كانت المراحل موجودة. الاستدعاء النهائي الحرج:

```
1,200/1,200 = 1,000
```

15.3. التحكم السلبي الحرج

الحالات الشبيهة بالحرجة لكنها ليست حرجة: تم العثور على حالات التحكم السلبي .600 مرشح حرج كاذب في الجولة الأولى: حدثت حالات 5. معدل الحرِج الكاذب في الجولة الأولى:

```
5/600 = 0.0083 = 0.83%
```

خمس حالات في المراجعة النهائية:

- ثلاثة رئيسية،
- اثنان متوسطة

تم تصحيحها كما هي. معدل الحرِج الكاذب النهائي: .0 الانعكاس النهائي للبوابة: حالات .0

16. المعنى الحقيقي للنتيجة الحرِجة

لا يعني الاستدعاء الحرج النهائي: 100% "يمكن لمقرر واحد اكتشاف جميع الأخطاء الحرجة." النتيجة التركيبية الحقيقية هي:

> تم اكتشاف جميع حالات التحدي الحرجة عند استخدام المراجعة المزدوجة، ومراجعة الخبراء، والمراجعة العليا معًا.

لذلك، لأسباب تتعلق بالتكلفة أو السرعة:

- المقرر الثاني،
- مراجع التخطي،
- مراجع الخبراء

تمت إزالتهم، ولا يمكن توقع نفس النتيجة. طريقة الاختبار، وليس الصيغة فقط:

تظهر أن طبقات الحوكمة الإلزامية

يجب أن تكون قابلة للاختبار أيضًا؛ هذا ليس تحققًا تجريبيًا.

17. نتيجة التصنيف الرئيسي

من الحالات المؤكدة 1,200 الرئيسية:

- الأغلبية هي حالات رئيسية حقيقية،
- جزء محدود منها معتدلة،

- بعضها مرشح لحالات حرجة،
- وبعضها غير محدد

في الجولة الأولى.

```
النهائي للإصدار الكبير F1:
```

كان 0.952. عتبة المرشح: 0.95 النتيجة:

تم تلبية الحد الأدنى للمرشح

لكن العتبة وحدها: تم تجاوزها بفارق 0.002 فقط. لذلك التصنيف الرئيسي:

- قوي،
- لكنه لا يحمل هامش أمان مريح

يجب تسجيله كحقل.

18. حالة الاستجابة: ارتباك

```
حالة RP الماكرو-F1:
```

لقد أصبح 0.928. اللبس الرئيسي:

الفئة الحقيقية	أكثر الاسترداد الخاطئ شيوعًا
RP-3 اجتياز مشروط	RP-4 إخفاق جسيم
RP-4 إخفاق جسيم	RP-3 اجتياز مشروط
RP-6 لم يتم حله	RP-4 إخفاق جسيم
RP-7 لا توجد استجابة قابلة للاستخدام	RP-3 اجتياز مشروط
RP-8 غير قابل للتقييم	RP-7 لا توجد استجابة قابلة للاستخدام

الفئة الحرجة: تم التعافي بالكامل في المرحلة النهائية. أعقد حد:

مع اجتياز مشروط و إخفاق جسيم

لقد تشكل بين. هذا الصعوبة مباشرة:

- الأهمية التراكمية،
- الإغفال المادي،
- إلى أي مدى تتعطل مهمتك الرئيسية

لقد نشأت من أسئلتك.

19. نتيجة تكامل الالتقاط

تم تصنيف حالات سلامة الالتقاط بشكل صحيح من :4,000 دقة صلاحية الالتقاط :3,928

```
3,928/4,000 = 0.982
```

عتبة المرشح: نتيجة :0.98

تم تلبية الحد الأدنى للمرشح

19.1. أخطاء الالتقاط

توزيع الحالات المصنفة خطأ من :72

الحالة الفعلية	قرار خاطئ	عدد الحالات
صالح حسب الشرط	صالح	20
بديل وصول صالح	دليل جزئي	8
خارج الموجة	صالح حسب الشرط	9
يُشتبه في التلاعب	دليل جزئي	12
نسخة شبه مكررة	صالح	5
صالح	صالح حسب الشرط	18
الإجمالي		72

لم يتم تصنيع أي عبوة أو التلاعب بها: لم يُؤخذ أي منها في التحليل النهائي كصحيح تمامًا. الاستجابة الدلالية:

- أن يُجبر على الانتماء إلى فئات

صحيح،

- خاطئ،
- حرج

لم يُرتبط به. هذا ليس نتيجة اختبار دلالة حقيقي؛ إنه افتراض لتصميم المثال الاصطناعي. هذا يوضح أن مبدأ الفصل بين الالتقاط والمعنى يعمل في العرض التوضيحي.

20. نتيجة الالتقاط في مجموعة 30,000 الرئيسية

داخل المجموعة الرئيسية:

مستوى الالتقاط	الملاحظة
NCL-4 مُجهزة وموقعة	27,850
NCL-3 حزمة الأدلة الكاملة	2,070
RP-8/ غير قابل للتقييم	80
الإجمالي	30,000

سبب سجلات 80 غير قابل للتقييم:

السبب	الملاحظة
عدم تطابق المُدخل/المخرجات	24
استجابة غير مكتملة	18
وقت خارج الموجة وغير موثوق	16

السبب	الملاحظة
دليل مكرر	12
تحويل نظام لا يمكن حله	10
الإجمالي	80

هذه السجلات :80

- لم يتم إعادة تصنيفها ك
- إجابة صحيحة,
- خطأ في المنتج,

إجابة غير صحيحة.

في التوزيع الرئيسي:

RP-8— غير قابل للتقييم

تم الاحتفاظ به كما هو.

21. استعادة الكيان

في النطاق الأساسي – حل الكيان: تم تقييم الردود الرئيسية لحالات 30,000 وحالات تصادم الكيان التشخيصي معًا. دقة حل الكيان: .0.987 الأخطاء الرئيسية:

- دفع الشركات التابعة يُحتسب ككيان رئيسي
- تُحتسب الامتيازات على أنها متجر يُدار مباشرة
- يُحتسب الكيان الهدف لتصادم الاسم غير ذي الصلة Apple-SYNTH
- اعتبر عائلة المنتج ككيان قانوني
- احتساب الشريك التاريخي كشريك تابع حالي

ليست جميع حالات الكيانات المزيفة على نفس مستوى الأهمية. تم العثور على حالات نقل السلطة الحرجة في مرحلة البوابة النهائية.

22. استرداد الدعم الواقعي

```
الحالة الواقعية ماكرو-F1:
```

0.923 أقوى الفئات:

- مدعوم
- تم التناقض
- الادعاء الرسمية مُنسبة بدقة

أضعف الفئات:

- غير مدعومة
- فجوة مرجعية

- مدعومة مع المؤهلات المطلوبة

الحد الذي كان صعبًا بشكل خاص هو: "لم يتم العثور على دليل" مقابل: "الأدلة المتاحة تتعارض مع الادعاء."
بعض المحكمين في وضع العالم المفتوح:

- غير مدعوم،
- متناقض

قد قاموا بعمل التمييز بشكل صارم للغاية. المراجعة النهائية صححت معظم هذه الحالات.

23. استرجاع النطاق

```
نطاق F1-Macro:
```

تم تجاوز عتبة المرشح .0.906 ومع ذلك، نظرًا لفشل جانب الإغفال في نظام النطاق والحدود، فإن هذه النتيجة وحدها غير كافية. النظام قوي عندما تُكتب ادعاءات النطاق بوضوح:

- "نفس السعر في جميع الدول."
- "الشركة الأم مرخصة في جميع الأسواق."

الصعوبة هي خطأ النطاق في:

- التوصية،
- الاسم القصير،
- الضمير،
- الحذف

زاد في الحالات حيث يكون مخفيًا.

24. الاسترجاع المتعلق بالنسبية والأسلوبية

```
النسبة الكلية (F1-Macro):
```

0.918

```
الأسلوبية الكلية (F1-Macro):
```

0.934 الحالات المنفصلة بشكل قوي:

- "تقول الشركة."
- "السجل الرسمي يظهر."
- "يدعي المستخدمون."
- "هناك قرار نهائي."
- "ربما."
- "غير موثق."

حالات النسبة الضعيفة المتبقية:

- عرض رأي موظف كموافقة من الجامعة،
- اعتبار المحتوى التحريري الممول كبحث مستقل،

عرض عناوين URL متعددة مشتقة كإجماع. هذه الحالات مرتبطة بضعف رسم خريطة الاستشهاد.

25. استرداد المكون

قيم الاسترداد لعشرة مكونات من مولد الحقيقة و NOMOS:

المكون	المولد	الاسترداد	خطأ مطلق
ENT	96.5	96.1	0.4
FAC	91.8	91.0	0.8
SBI	84.6	82.0	2.6
TLA	88.7	88.0	0.7
ECI	82.1	78.7	3.4
EPI	87.9	85.8	2.1
TCR	87.0	88.2	1.2
LGF	82.4	84.8	2.4
USI	85.2	86.9	1.7
STR	87.0	90.7	3.7
إجمالي NOMOS	873.2	872.2	1.0

متوسط الخطأ المطلق للمكون:

```
MAE_component= 1.9
```

الحد الأدنى للمرشح:

```
≤2.5
```

النتيجة:

تم تلبية الحد الأدنى للمرشح

ومع ذلك، تحتوي هذه الجدول أيضًا على تحذير هام. يتم حساب ECI منخفضًا بمقدار 3.4 نقطة، وSBI منخفضًا بمقدار 2.6 نقطة، وSTR مرتفعًا بمقدار 3.7 نقطة، وLGF مرتفعًا بمقدار 2.4 نقطة. بعض هذه الأخطاء ألغت بعضها البعض، وانحرف المجموع المركب بمقدار نقطة واحدة فقط. لذلك:

> المظهر القوي للتعافي المركب لا يزيل متطلبات الإبلاغ عن تعافي المكونات.

26. سوء تفسير التعافي المركب

لا يمكن استخدام هذه الجملة بمفردها: "استعاد NOMOS المجموع المركب الاصطناعي المحدد في المُولّد بفارق نقطة." الجملة الصحيحة هي:

> استعاد *NOMOS* النتيجة المركبة بفارق نقطة واحدة؛ ومع ذلك، وقعت أخطاء اتجاهية أكبر في مكونات ECI وSBI وLGF وSTR، مما أدى جزئيًا إلى تعويض بعضها البعض.

مهمة المعيار ليست فقط إيجاد المجموع الصحيح. بل هي شرح السبب وراء حدوث المجموع بشكل صحيح.

27. استعادة العدالة في اللغة والجغرافيا

تم تحديد قواعد اللغة والبلدان الاصطناعية سابقًا ضمن Generator Truth. خطأ الاستعادة المطلق لمكون LGF: 2.4 نقاط. حد المرشح:

```
≤3
```

النتيجة:

تم تلبية الحد الأدنى للمرشح

27.1. مقاييس فرعية

معدل العدالة	المولد	الاسترداد	الخطأ
قاعدة اللغة Q1 0	68.0	69.7	1.7
الفرق بين اللغة Q9 Q1–0 0	25.0	22.8	2.2
القاعدة الجغرافية Q1 0	72.0	71.1	0.9
الفرق الجغرافي	18.0	16.4	1.6
عامل التغطية	0.94	0.95	0.01
LGF	82.4	84.8	2.4

الاسترداد:

- تم تقدير أدنى أداء للغة بشكل مبالغ فيه قليلاً،
- تم تقليل تقدير التفاوت بشكل مبالغ فيه قليلاً

أظهر هذا أن حالة الإنصاف أفضل قليلاً مما هي عليه في الواقع.

28. التمييز بين خلل المُدخل وخلل لغة الذكاء الاصطناعي

داخل ملحق التكافؤ اللغوي:

- 500 فشل لغة الملف الشخصي الحقيقي للذكاء الاصطناعي،
- 500 فشل تكافؤ المُدخل

تم العثور على حالة. التصنيف الصحيح للسبب:

```
938/1,000 = 0.938
```

لقد حدث ذلك. من الحالات المصنفة خطأً في :62

- خلل في المُدخل في منتج الذكاء الاصطناعي th 39،
- خلل في لغة الذكاء الاصطناعي في المُدخل على rd 23

تم تحميله. هذا الخطأ أثر مباشرة على LGF. مر مرشح مكون العدالة العتبة. ومع ذلك:

> لم يصل تصنيف المسؤولية السببية بعد إلى مستوى الأمان الخارجي.

29. استعادة اللوحة النظيفة-الطبيعية

متوسط الفرق في اللوحة لحقيقة المولد في مطابقات 1,000 النظيفة—الطبيعية ضمن الملحق التشخيصي:

G_{N−C}^G = −8.4

نقاط. استعادة NOMOS:

G_{N−C}^N = −7.9

نقاط. الفرق المطلق: 0.5 نقاط.

```
متوسط الخطأ المطلق لفجوة لوحة الملف الشخصي الفردي:
```

هو 1.6 نقاط.

29.1. الانحراف في التخصيص الشخصي

هناك 173 حالات انحراف مادي في التخصيص الشخصي في Generator Truth. تم العثور عليها: 166 استرجاع:

```
166/173 = 0.9595
```

من بين الحالات السبعة الفائتة:

- أربعة هي إغفال الحدود،
- اثنتان تعليمات مواتية للكيان.
- واحدة هي تسريب بيانات مقيد.

الحالة. تم العثور على تسريب بيانات مقيدة في المراجعة النهائية للخصوصية وفتح بوابة حرجة. تظهر هذه النتيجة مرة أخرى أهمية التحكيم متعدد الطبقات.

30. الموجة واستعادة الاستقرار

في ثلاث موجات اصطناعية، أُدخلت التغييرات الآتية: في W2، تغير وضع الويب والاستشهاد لأحد المنتجات. ووقع حدث خارجي اصطناعي في W3. أنتج SYNTH-AI-10 حدثًا حرجًا في W2 ثم إصلاحًا في W3. وتراجع SYNTH-AI-03 في الأداء باللغات محدودة الموارد. وحمل SYNTH-AI-07 بيانات قديمة عن الأسعار وتغطية المنتجات. NOMOS:

- قسم بشكل صحيح التغيير داخل الموجة للنظام إلى موجتين فرعيتين
- فصل إصدارات حزمة الحقيقة قبل الحدث الخارجي وبعده،
- تم تحديث الوضع الحالي بعد التصحيح،

تم الحفاظ على الحدث الحرج التاريخي.

30.1. الإفراط STR في التقدير

مولد STR: 87.0 استعادة STR: 90.7 خطأ: +3.7 نقاط. السبب: بعض التغيرات داخل الموجة تم فصلها كإصدارات صريحة بدلاً من الانحراف النظامي، وصيغة التقييم تمنح مكافأة مفرطة على الاستقرار داخل الموجات الفرعية المفصولة. تشير هذه النتيجة إلى أن صيغة STR، بعد التغيير، لا تميز بشكل كافٍ بين الاستقرار والاستقرار النظامي المتواصل.

31. اختبار الحدث الحرج النادر

تم حساب أربعة سيناريوهات منفصلة للأحداث النادرة الاصطناعية؛ هذه ليست تجارب ميدانية أو برمجية. النتائج التالية هي مجرد عرض للطريقة فقط:

عدد الأحداث الحرجة الافتراضي	معدل تقريبي موزون / 1,000	حد الثقة العلوي أحادي الجانب لـ 95%/ 1,000	سلوك مراجعة الأحداث

عدد الأحداث الحرجة الافتراضي	معدل تقريبي موزون / 1,000	حد الثقة العلوي أحادي الجانب لـ 95%/ 1,000	سلوك مراجعة الأحداث
0	0.00	2,991	لم يتم إثبات ادعاء المخاطر الصفرية
1	1.08	4,735	تم فتح سجل CIS-2
2	2.15	6,282	مراجعة من قبل المدقق
5	5.38	10,484	التكرار على مستوى الخلية

ملاحظة: الحدود العليا هي حدود الثقة العليا أحادية الجانب لطريقة كلوبّر–بيرسون *95%* تحت افتراض وجود n=*1,000* من الملاحظات المستقلة بنرّولي. لا يمكن تطبيقها مباشرة على تصميمات العينات المرجحة أو المعقدة [K09].

في سيناريو الحدث الصفري، لم يقل النظام: "المخاطرة الحرجة صفر." الإخراج الصحيح:

«الحدث الحرج المؤكد المرصود = `0`؛ وبافتراض `n=1,000` ملاحظة مستقلة، يبلغ حد الثقة الأعلى أحادي الطرف بطريقة `Clopper-Pearson` عند مستوى `95%` مقدار `2,991/1,000`».

قد حدث. سلوك الحدث النادر المحتمل قد استوفى العتبات الاصطناعية المحددة.

32. معدل تغطية فترة الثقة

عملية توليد السكان الاصطناعي: تم تقديم مثال لطريقة تم حسابها على عينات افتراضية .2,000 تم حساب معدلات التغطية للفواصل الاسمية :95%

التغطية التجريبية	المؤشر
94.6%	النجاح الصارم
95.0%	النجاح المقبول
94.1%	مركب NOMOS
93.7%	LGF
95.2%	الحد الأعلى لمعدل النقدي
94.8%	FAC
93.5%	SBI

نطاق قبول المرشح: محدد كـ .93%–97% جميع معدلات التغطية الأساسية تقع ضمن نطاق المرشح. النتيجة:

تم تلبية الحد الأدنى للمرشح

33. اختبارات التحول

الإجمالي: تم استخدام 3,200 أزواج استجابة متحولة.

33.1. التحويلات المحافظة على المعنى

إعادة صياغة ترتيب الجملة، تنسيق آمن، موضع الاستشهاد، طالما أن الرابط يبقى.
الكيان المكافئ الاسم المستعار المتوقع: يجب ألا تتغير حالة مراجعة النظراء والاستجابة. النتيجة الثابتة الدقيقة:

33.2. التحولات التي تغير المعنى

إزالة الإسناد زيادة اليقين النقل إلى الكيان غير الصحيح تغيير الوقت الحالي من التاريخي إزالة جملة الحدود إضافة ادعاء مصدر مستقل المتوقع: يجب أن يتغير البعد القرار ذي الصلة. الحساسية الصحيحة:

33.3. الفوالق الميتامورفية الكبرى

نقل الاقتباس من الجملة إلى نهاية الفقرة، التقييم غير المتسق لصيغة 'يبدو' و 'من المحتمل'، تفسير اسم العلامة التجارية البديل ككيان قانوني، جملة الحد الفاصل موجودة في فقرة أخرى—تتوافق هذه النتائج مع نتائج الاقتباس والحذف.

34. نتائج الاختبار المضاد

حالة الفحص المضاد	تصرف NOMOS الصحيح
طليق جدًا ولكنه غير صحيح جوهريًا	98.8%
صحيح جدًا، خطأ واحد فقط في الذرة	رد 100% فشل تلقائي
إجابة قصيرة لكنها كافية	99.2% اجتياز صحيح
إخلاء مسؤولية عامة وخطأ ملموس	تم حفظ 98.7% بشكل غير صحيح
غسل الأدلة	91.8% صحيح
نقل ترخيص الكيان غير الصحيح	99.4% صحيح
تجاوز حد التوصية	82.9% صحيح
تصديق جامعة مزورة	96.1% صحيح
الرقم صحيح، الاستقلالية خاطئة	92.0% صحيح
درجة مناسبة من عدم اليقين	97.3% صحيح

أضعف فئة اختبار مضاد:

تجاوز حد التوصية

قد كانت.

`تؤكد هذه النتيجة مرة أخرى فشل Material Omission F1.`

35. تم العثور على أخطاء في الطريقة بواسطة الاختبار

في نهاية العرض التوضيحي الاصطناعي، تم تسجيل خمسة نتائج طرقية لـ NOMOS 0.9.

نتيجة الطريقة 01

غموض إرفاق الاستشهاد

`الخطورة: كبيرة / المكون المتأثر: ECI / النتيجة: لم يتم اجتياز حد F1 لاستشهاد المرشح.`

القاعدة المحددة لتحديد الجمل التي تدعم الاستشهاد على مستوى الفقرة غير دقيقة بما فيه الكفاية.

نتيجة الطريقة 02

عدم اكتشاف تخطي حدود المواد

`الأهمية: كبيرة / المكونات المتأثرة: TCR، SBI، البوابة الحرجة / النتيجة: لم يتم تحقيق حد F1 لتخطي المواد.`

الحدود التي لم تُذكر صراحة في جمل التوصية والعلاقة ولكنها تغير قرار المستخدم لم يتم التقاطها بشكل كافٍ.

نتيجة الطريقة 03

خطأ في الإسناد / خطأ في الذكاء الاصطناعي

الأهمية: معتدلة / المكون المتأثر: LGF / النتيجة: على الرغم من تجاوز درجة العدالة، فإن تعيين المسؤولية خاطئ.

نتيجة الطريقة 04

المكافأة المفرطة على الاستقرار بعد تقسيم الموجة الفرعية

الأهمية: معتدلة / المكون المتأثر: STR / النتيجة: تم احتساب STR نقاطًا أعلى بمقدار 3.7 من الحقيقة الناتجة.

نتيجة الطريقة 05

القبول البديل لإمكانية الوصول أقل من اللازم

الأهمية: متوسطة / الطبقة المتأثرة: سلامة الالتقاط / النتيجة: تم تقليص ثمانية حالات صحيحة للقبول على مستوى الوصول إلى مستوى أدلة غير كافٍ.

36. لماذا هذا ليس عرضًا توضيحيًا BQ-3؟

بالنسبة لـ BQ-3 — تم اجتياز التحقق المختوم:

يجب تلبية جميع الحدود القصوى للمرشحين الإلزاميين، ويجب ألا توجد انتهاكات لا تُحتمل. لم يتم تلبية حدين ماديين:

المعيار	النتيجة	الحد
معامل F1 لرسم الخرائط الاستشهادية	0.887	≥ 0.90
معامل F1 للإغفال المادي	0.824	≥ 0.85

بالإضافة إلى ذلك:

- لم يتم إعادة إنتاج الفريق المستقل،

- ولم يتم نشر النص الحقيقي والرمز.

لم يتم تنفيذ تجديد الاحتفاظ غير الملموس. لذلك، البيان التصويري الصحيح هو:

DEMO-BQ-2 — اكتملت تجربة خط المعالجة الجاف؛ مطلوب تصحيح المواد

يجب أن يكون كذلك. الحالة الفعلية للاختبار هي:

BQ-0 — لم يتم التنفيذ

تبقى كما هي.

37. لماذا لم يتم خفض العتبات؟

نتيجة رسم الخرائط المرجعي: 0.887 العتبة المقترحة: 0.90 الفرق صغير. لو كنا قد حددنا العتبة على 0.885، لكان الأسلوب قد اجتاز. نتيجة إغفال المواد: 0.824 العتبة المقترحة: 0.85 لو كنا قد حددنا العتبة على 0.82، لكان هذا الحقل قد اجتاز أيضًا. ومع ذلك، هذه التغييرات:

- كانت ستتم بعد رؤية النتائج.
- من أجل ضمان اجتياز NOMOS

هذا هو أدنى انتهاك أساسي للاختبار. السلوك الصحيح هو:

- الحفاظ على العتبة
- فشل النشر
- تصحيح الطريقة
- إنشاء نسخة جديدة
- إعادة الاختبار على العينة غير المعدلة

يجب أن تكون.

38. خطة تصحيح الطريقة 1.0

38.1. مُرَسِم الإشارات 1.0

النظام الجديد:

- النطاق المصدر للإشارة,
- حقل إرفاق الادعاء,
- نطاق داخل الفقرة,
- حالة الإسناد فقط،
- عائلة نسب المصدر

ستكون إلزامية. لكل اقتباس:

```
أي الذرات بالضبط يدعمها هذا المصدر؟
```

سيتم تحويل السؤال إلى سجل مفتوح.

38.2. دليل إهمال الحدود 1.0

لكل توصية ومعاملة مُدخلة:

- الحدود الإلزامية التي تغير قرار المستخدم،
- اختبار القرار الافتراضي،
- مصفوفة الكيان–الخدمة–الاختصاص

ستكون محددة مسبقًا. اختبار افتراضي:

```
إذا تم إبلاغ المستخدم بهذا الحد، هل ستتغير توصية المستخدم المعقولة أو قرار المعاملة؟
```

إذا كانت الإجابة نعم، فإن السهو سيخضع على الأقل لمراجعة موضوعية.

38.3. الحكم على مسؤولية الذكاء الاصطناعي

إسقاط اللغة:

- تكافؤ التحفيز,
- سلوك منتج الذكاء الاصطناعي،
- الالتقاط,
- موقع المستخدم

سيتم إنشاء متجه قرار منفصل لتحديد أي من المصادر ينتمي إليه.

38.4. مراجعة صيغة STR

STR الجديد:

- الاستقرار ضمن الموجة الفرعية،
- تغيير إصدار النظام،
- الاستقرار بعد التصحيح,
- سلسلة زمنية متتابعة بدون انقطاع

سوف تفصل الحقول إلى مكونات فردية.

38.5. معادلة أدلة الوصول

غير مرئي ولكن يحمل أدلة معادلة:

- نص مكتوب،
- شجرة الوصول،
- سجل تكنولوجيا المساعدة،
- تسجيل جلسة صوتية

سيتم إنشاء مصفوفة تعادل مفتوحة من أجل.

39. قاعدة إعادة الاختبار

طريقة 1.0: لا يمكن تعديلها مرارًا على نفس مجموعة التحقق المختومة. مطلوب اختباران:

A. مجموعة الانحدار

يتضمن الحالات الفاشلة لطريقة 0.9. الغرض: لمشاهدة أن الإصلاح يعمل بالفعل.

ب. الاستبقاء التجديدي غير المعدل

يشمل الحالات التي لم تُر أبدًا عند تصميم طريقة 1.0. الغرض: للتحقق من عدم حدوث التكيف الزائد مع الاختبار القديم. فقط BQ-3:

- نجاح الانحدار،
- نجاح الاستبقاء غير المعدل

يمكن تقديمه إذا تم الحصول عليه معًا.

40. قرار الانتقال إلى العالم الحقيقي

استنادًا إلى هذا العرض التركيبي، القرار الصحيح لـ NOMOS :0.9

> غير مستعد لمنح الشركات الحقيقية درجة أهلية متاحة للجمهور أو علامة *950+. NOMOS*

الأسباب:

- مرشح خريطة المرجع تحت الحد الأدنى
- كشف الإغفال المادي للمرشح تحت الحد الأدنى
- لا يوجد إعادة إنتاج مستقلة
- لم يتم إجراء اختبار حقيقي
- مستوى BQ-4 لم يتم الوصول إليه

ومع ذلك، الطريقة:

- استعادة التوزيع،
- البوابة الحرجة،
- حالة الاستجابة،
- استعادة المكون،
- فاصل الثقة،
- سلوك الحوادث النادرة

يُظهر أساسًا قويًا من حيث الصيانة. الحالة الصحيحة:

> جاهز للبحث والاستخدام التجريبي؛ غير جاهز بعد للمعيار العام وعلامة المطابقة.

41. بطاقة قبول الاختبار الاصطناعي

المعيار	النتيجة	الحكم
حدود الادعاء F1	0.964	اجتاز
دقة الكيان	0.987	اجتاز
F1 الكلي للحالة الواقعية	0.923	اجتاز
F1 الكلي للنطاق	0.906	اجتاز
F1 الكلي للإسناد	0.918	اجتاز
معامل F1 لرسم الخرائط الاستشهادية	0.887	لم يجتز

المعيار	النتيجة	الحكم
معامل F1 للإغفال المادي	0.824	لم يجتز
الاستدعاء النهائي للحالات الحرجة	1,000	اجتاز
الإيجابيات الكاذبة النهائية للحالات الحرجة	0	اجتاز
F1 الكلي للحالات الجسيمة	0.952	اجتاز
F1 الكلي لـ RP	0.928	اجتاز
متوسط الخطأ للمكوِّن	1.9	اجتاز
الخطأ المركب	1/1,000	اجتاز
خطأ العدالة	2.4	اجتاز
دقة الالتقاط	0.982	اجتاز
تغطية CI	ضمن 92%–98%	اجتاز
عكس البوابة	0	اجتاز
التكاثر المستقل	لا شيء	غير مكتمل
تنفيذ مجموعة البيانات الفعلية	لا شيء	غير مكتمل

42. القاعدة المشتركة للعرض

هذا التشغيل الاصطناعي يوضح ثلاثة أشياء في نفس الوقت عن NOMOS .0.9

42.1. المناطق التي يكون فيها NOMOS قويًا

استعاد توزيع حالة الاستجابة بدقة عالية. لم تفشل سلسلة الحوكمة النهائية في أي بوابة حرجة ولم تخف حدثًا حرجيًا ضمن متوسط مرتفع. ميز بفعالية بين أخطاء الكيان الخطأ والأخطاء الزمنية وأخطاء دعم الحقائق. لم يحوّل سجلات غير قابل للتقييم إلى عيوب في المنتج، ولم يستنتج صفر مخاطر من عدم وجود أحداث حرجة ملحوظة، وحافظ على التمييز بين الحوادث التاريخية والتصحيح الحالي، واستعاد الدرجة المركبة بخطأ مطلق منخفض.

42.2. المجالات التي يكون فيها NOMOS ضعيفًا

لم يتمكن من تحديد أي الادعاءات يدعمها الاقتباس بشكل كافٍ. لم يتمكن من اكتشاف الإغفالات المهمة في التوصية بشكل كافٍ. لم يميز دائمًا بشكل صحيح بين عيوب المُدخل وعيوب لغة الذكاء الاصطناعي. بعد التمييز بين الموجات الفرعية، بالغ قليلاً في مكافأة الاستقرار. اعتبر بعض الأدلة المتعلقة بإمكانية الوصول أضعف من اللازم.

42.3. الأشياء NOMOS لم تثبت بعد

لم يثبت بعد أنه يعمل على المستخدمين الحقيقيين، أنه يمكن تطبيقه على منتجات الذكاء الاصطناعي الحقيقية، أنه يمتلك تعادلاً حقيقيًا في التعليمات متعددة اللغات، أن الجامعات المستقلة ستنتج نفس النتيجة، أن الحوكمة طويلة المدى ستعتمد على اعتبارات تعارض المصالح، أو أن علامة الامتثال العامة يمكن منحها قانونيًا بثقة.

43. الأحكام الإلزامية المعيارية

CH18-N01

لا يمكن تقديم العرض التوضيحي الاصطناعي في الكتاب كما لو كان جولة اختبار عامة حقيقية.

CH18-N02

يجب فصل حالة العرض التوضيحي الاصطناعي عن حالة اختبار BQ الحقيقية.

CH18-N03

يجب أن تحافظ الاختبار على حالة BQ-0— غير منفذ حتى يتم إجراء مجموعة بيانات حقيقية وتشغيل مستقل.

CH18-N04

لا يمكن نسب نتائج العرض إلى أداء جهاز Apple حقيقي أو مزود ذكاء اصطناعي حقيقي.

CH18-N05

يجب الاحتفاظ بالاستجابات الرئيسية 30,000 والمجموعات التشخيصية والأهمية والالتقاط والذهب في صناديق منفصلة.

CH18-N06

لا يمكن لحالات التحدي تغيير توزيع الانتشار الأساسي.

CH18-N07

يجب تثبيت توزيع حقيقة المولد قبل نتائج الاسترداد.

CH18-N08

قبل فتح النتائج الحقائقية للمولد، يجب تثبيت استجابة NOMOS ونتائج الدرجات.

CH18-N09

لا يمكن تقييم التعافي بناءً على النتيجة المركبة وحدها.

CH18-N10

يجب نشر حدود الادعاء، الكيان، الواقعية، النطاق، الوقت، الإسناد، الاستشهاد، السهو، مستوى الأهمية، الاستجابة، المكون، والاسترداد المركب بشكل منفصل.

CH18-N11

لا يمكن لأخطاء الدرجة المركبة إخفاء أخطاء المكونات لمجرد أنها صغيرة.

CH18-N12

لا يمكن احتساب أخطاء المكونات التي تلغي بعضها البعض حسابيًا على أنها استعادة سببية ناجحة.

CH18-N13

يجب مقارنة توزيع الاسترداد من RP-1 إلى RP-8 بشكل صريح مع توزيع الحقيقة للمولد.

CH18-N14

يجب عرض عمليات الاسترداد للنجاح الصارم والنجاح المقبول بشكل منفصل.

CH18-N15

يجب الإبلاغ عن عدد الاستجابات الحرجة واسترداد البوابة الحرجة بشكل منفصل.

CH18-N16

يجب أن يُحتفظ بفروق الاستدعاء الحرج الأولية للمحكم وفروق الاستدعاء الحرج المحكّم النهائي منفصلة.

CH18-N17

يجب أن يكون اعتماد الاستدعاء الحرج النهائي على طبقات الحوكمة مرئيًا.

CH18-N18

لا يمكن افتراض نفس النتيجة عندما يصدر حكم ثانٍ أو مراجعة خبراء.

CH18-N19

يجب نشر معدل الإيجابيات الكاذبة الحرجة جنبًا إلى جنب مع الاسترجاع.

CH18-N20

ينبغي أن تكون الضوابط السلبية المشابهة للحرجة جزءًا إلزاميًا من نظام تقييم الأهمية النهائي.

CH18-N21

إذا كان عتبة المرشح F1 لمطابقة المرجع أقل من ذلك، فلا يمكن إعلان الاختبار ناجحًا.

CH18-N22

إذا كان المرشح لحذف المادة F1 أقل من العتبة، فلا يمكن الإعلان عن أن الاختبار ناجح.

CH18-N23

لا يمكن خفض عتبات المرشح بعد رؤية نتائج التعافي.

CH18-N24

تتطلب تغييرات العتبة إصدار طريقة اختبار جديدة والتحقق من صحة جديدة غير ملامسة.

CH18-N25

لا يمكن أن يكون الفرق الطفيف في العتبة مبررًا لإخفاء الفشل.

CH18-N26

يجب أن يكون حذف حدود المادة في التوصية قابلاً للتدقيق تمامًا مثل ادعاء خاطئ واضح.

CH18-N27

لا يمكن لموقع المرجع داخل الفقرة أن يخلق تلقائيًا دعمًا كاملًا للفقرة.

CH18-N28

لا يمكن احتساب التعافي للاقتباسات المخصصة للنسبة فقط أو للتحقق الواقعي من الحقائق.

CH18-N29

يجب أن تظهر أخطاء النسب المصدر أيضًا في استرداد رسم الخرائط الاستشهادية.

CH18-N30

يجب حساب دقة الالتقاط بشكل مستقل عن صحة الاستجابة الدلالية.

CH18-N31

الالتقاط المزيف الذي يحمل الإجابة الدلالية الصحيحة لا يجعل الالتقاط صالحًا.

CH18-N32

لا يمكن احتساب استجابة صحيحة لكنها خاطئة كخطأ في الالتقاط.

CH18-N33

لا يتعين أن يكون معيار الأدلة البديلة للولوج في نفس شكل المرئي.

CH18-N34

عندما يتم تقليل دليل الوصول الصالح دون داعٍ إلى مستوى أدنى، يجب فتح نتيجة طريقة.

CH18-N35

يجب أن تظل غير قابل للتقييم مرئية في توزيع الاستجابة الرئيسي.

CH18-N36

لا يمكن إزالة غير قابل للتقييم من المقام لزيادة استرداد المشاهدات.

CH18-N37

يجب أن تحمل عيوب صياغات المُدخل وعيوب لغة الذكاء الاصطناعي فئات سبب منفصلة.

CH18-N38

لا يمكن تقييم استرداد العدالة بناءً على فرق المركب LGF فقط.

CH18-N39

يجب أيضًا نشر خط الأساس اللغوي والفروق والتغطية وتعيين المسؤولية.

CH18-N40

لا ينبغي لاسترداد لوحة نظيفة–طبيعية أن ينتج بند تخصيص سببي.

CH18-N41

يجب أن تحمل حالات انحراف التخصيص مقاييس منفصلة للاستدعاء والإيجابيات الكاذبة.

CH18-N42

لا يمكن تقليل تسرب البيانات المقيدة باعتباره نتيجة تخصيص طبيعية.

CH18-N43

يجب أن يُنتج تغيير النظام داخل الموجة بنية فرعية صحيحة للموجة.

CH18-N44

إذا كانت فصل الموجة الفرعية يزيد بشكل مصطنع درجة الاستقرار، يجب الكشف عن صيغة الاكتشاف.

CH18-N45

لا يمكن للتصحيح الحالي محو الحدث الحرج التاريخي.

CH18-N46

يجب أن تتضمن اختبارات الأحداث النادرة سيناريوهات صفرية وواحدة ومتعددة للأحداث.

CH18-N47

لا يمكن لسيناريو حرج صفر ملاحظ أن ينتج نتيجة مخاطر صفرية.

CH18-N48

لا يمكن لحدث حرج واحد أن يحدد، بمفرده، معدل فشل نظامي 100 بالمئة.

CH18-N49

يجب اختبار طريقة فترة الثقة مع التغطية التجريبية في السكان الاصطناعيين المتكررين.

CH18-N50

يجب نشر النتيجة الفعلية للتغطية لفترات النسبة المئوية الاسمية 95.

CH18-N51

يجب حساب اختبارات الثوابت الميتامورفية والاختبارات الحساسة للتوقعات بشكل منفصل.

CH18-N52

يجب ألا يغير إعادة الصياغة التي تحافظ على المعنى نتيجة المراجعة بلا داعٍ.

CH18-N53

يجب أن ينتج عن التحويل الذي يغير معنى النسبة أو الكيان أو الوقت أو الحدود تغييرًا مناسبًا في القرار.

CH18-N54

يجب أن تغطي اختبارات الاختبار المضاد بطلاقة فقط حالات حرجة، غسيل الأدلة، والإغفال.

CH18-N55

يجب أن تكون أضعف فئات الاختبار المعارضة مرئية في بطاقة الاختبار العامة.

CH18-N56

يجب تحويل نتائج الطريقة إلى سجلات منفصلة مع تعيين إصدارات لكل منها من نتائج الاختبار.

CH18-N57

يجب إعطاء أهمية نتائج الطريقة وفقًا لتأثير المستخدم الفعلي والتدقيق.

CH18-N58

لا يمكن تخفيض فشل المراجع والحذف إلى مستوى استشاري فقط بسبب اختلاف بسيط في المقياس.

CH18-N59

يجب تعيين مستوى جودة الاختبار وفقًا لجميع الحدود الإلزامية وشروط الاستقلال.

CH18-N60

إذا فشل حدان ماديان، لا يمكن منح BQ-3.-DEMO

CH18-N61

لا يمكن منح BQ-4 بدون إعادة إنتاج مستقلة.

CH18-N62

لا يمكن اعتبار نتيجة عرض داخل الكتاب BQ كحالة فعلية دون تنفيذ تجربة حقيقية.

CH18-N63

لا يمكن لتصحيح الطريقة تغيير نتيجة العرض القديم.

CH18-N64

يجب أن تحمل الطريقة 1.0 النسخة الجديدة من الدرجة وكتاب الشفرات.

CH18-N65

مجموعة الانحدار وحدها لا يمكنها التحقق من الطريقة الجديدة.

CH18-N66

يجب أن يكون الاحتفاظ المتجدد غير الممسوس إلزاميًا للطريقة الجديدة.

CH18-N67

لا يمكن تكييف الطريقة الجديدة مع نفس مجموعة التحقق مرارًا وتكرارًا.

CH18-N68

لا يمكن بدء التدقيق العام المباشر دون تحقيق الحد الأدنى من جودة الاختبار وشرط الاستنساخ المستقل.

CH18-N69

لا يمكن منح علامة المطابقة بناءً على نجاح العرض الاصطناعي.

CH18-N70

يجب الكشف عن المناطق التي فشلت فيها الاختبارات للجمهور.

CH18-N71

لا يمكن نشر العتبات التي يتم تجاوزها بمفردها.

CH18-N72

لا يمكن للنجاح المركب في الاسترداد أن يمحو نتيجة مرجع فاشل أو إغفال.

CH18-N73

يجب أن تكون نقاط القوة والضعف في العرض على نفس البطاقة العامة.

CH18-N74

يجب سرد الادعاءات التي لم يثبتها NOMOS بوضوح.

CH18-N75

لا يمكن تحويل الأرقام الاصطناعية داخل الكتاب إلى اقتباسات أو بيانات عامة كما لو كانت بيانات منفذة حقيقية.

CH18-N76

يجب أن تحمل بيانات العرض الاصطناعية أيضًا علامة اصطناعية بصيغة قابلة للقراءة آليًا.

CH18-N77

لا يمكن فهرسة مجموعة العرض على الإنترنت كمعلومات حقيقية للشركة.

CH18-N78

عندما يبدأ الاختبار الفعلي، يجب استخدام هوية وإصدار جديدين.

CH18-N79

جميع حسابات الاسترداد تتطلب رمزًا قابلاً لإعادة الإنتاج وملفًا توضيحيًا.

CH18-N80

يجب أن يكون لكل تشغيل اختبار، أو استرداد، أو إيجاد طريقة، أو تصحيح، أو قرار نشر، مالك بشري أو مؤسسي مسؤول.

44. أشكال الفشل

CH18-F01 — عرض العد باعتباره تشغيلًا حقيقيًا

الأرقام الموجودة في الكتاب معروضة كنتائج اختبار عامة.

CH18-F02 — دمج BQ-0 مع DEMO-BQ-2

يقال إن الاختبار تم التحقق من صحته دون تنفيذ فعلي.

CH18-F03 — عبادة الدرجات المركبة

يغطي خطأ مركب مقداره نقطة واحدة جميع الأخطاء الكامنة تحته.

CH18-F04 — اعتبار إلغاء المكون نجاحًا

تلغي الأخطاء الاتجاهية العالية والمنخفضة بعضها البعض.

CH18-F05 — نشر استرداد النجاح الصارم فقط

تختفي اختلاطات فئة RP.

CH18-F06 — احتساب الاستدعاء النهائي الحرج على أنه نجاح الحكم الأول

يصبح تأثير الخبير والمراجعة العليا غير مرئي.

CH18-F07 — اعتبار الحكم الثاني غير ضروري

يتم إخفاء السبعة الحرجة التي تم تفويتها في المرور الأول.

CH18-F08— تجاهل النقدي الزائف
يتم تقديم نظام الأهمية القصوى على أنه نجاح.
CH18-F09— خفض عتبة الاقتباس
تم تغيير العتبة بحيث يجتاز نتيجة 0.887.
CH18-F10— خفض عتبة الحذف
تم تغيير قاعدة الاختبار بحيث يجتاز نتيجة 0.824.
CH18-F11— النجاح عن طريق قول "فشل بفارق ضئيل جدًا"
تفقد ربط العتبة الذي كان مثبّتًا مسبقًا.
CH18-F12— نشر النسبة المئوية على الفقرة
تعتبر المصدر الواحد قد دعم جميع الادعاءات.
CH18-F13— احتساب مصدر يعتمد فقط على النسبة كمعلومة
ما تقوله الشركة يصبح حقيقة مستقلة.
CH18-F14— تجاهل نسب المصدر
تصبح روابط URL المشتقة دليلاً مستقلاً.
CH18-F15— اعتبار الإغفال تفصيلًا اختياريًا
يتم تقليل الحد الذي يغير قرار المستخدم.
CH18-F16— جعل كل إغفال أمرًا حرجًا
التفاصيل الاختيارية تنتج درجة مفرطة من الأهمية.
CH18-F17— اعتبار عيب صياغة المُدخل كخطأ من الذكاء الاصطناعي
يتم تعيين مسؤولية العدالة بشكل خاطئ.
CH18-F18— جاري رفع خطأ الذكاء الاصطناعي إلى المُطالبة
يتم نقل فساد اللغة الفعلي إلى أداة القياس.
CH18-F19— النظر في تكوين العدالة كافٍ
يتم الاحتفاظ بالقاعدة واسترداد الفجوة.
CH18-F20— تجاهل التنبؤ الزائد للسلسلة
الانفصال تحت الموجي يتحول إلى مكافأة على الاستقرار.
CH18-F21— معاقبة 'الوصولية' في دقة الالتقاط
يعتبر الدليل البديل المعادل ضعيفًا.
CH18-F22— احتساب 'غير قابل للتقييم' كفشل في المنتج
يتحول عيب الالتقاط إلى أداء الذكاء الاصطناعي.
CH18-F23— حذف 'غير قابل للتقييم'
تزداد التغطية القابلة للتقييم بشكل صناعي.
CH18-F24— عدم نشر الأخطاء المورفولوجية
يُحتفظ به رغم أن نفس المعنى يؤدي إلى قرار مختلف.
CH18-F25— إخفاء أضعف فئة في الاختبار المعاكس

نتيجة لإغفال الحدود، يتم إزالته من البطاقة العامة.

CH18-F26— فقط مقياس الحكم النهائي

تصبح نقطة الضعف في العملية الأولية غير مرئية.

CH18-F27— فقط مقياس المرور الأول

تصبح القدرة التصحيحية لسلسلة الحوكمة غير مرئية.

CH18-F28— الاختبار بدون إيجاد الطريقة

المناطق الفاشلة تبقى فقط كأرقام.

CH18-F29— جعل إيجاد الطريقة إرشاديًا

تأثيره على النتيجة المادية يقل.

CH18-F30— الإعلان عن BQ-3

لم يتم تجاوز عتبتين إلزاميتين.

CH18-F31— الإعلان عن BQ-4

لا يوجد إنتاج مستقل.

CH18-F32— عد العرض التوضيحي الاصطناعي كتحقق مباشر

لا يوجد مستخدم حقيقي أو منتج ذكاء اصطناعي.

CH18-F33— الإفراط في التكيف مع مجموعة الانحدار

طريقة 1.0 تحفظ فقط الأخطاء القديمة.

CH18-F34— عدم استخدام مجموعة احتياطية لم تُلمس

لم يتم اختبار التعميم.

CH18-F35— استبدال العرض التجريبي القديم بصمت بصيغة جديدة

تاريخ الإصدارات مفقود.

CH18-F36— منح شارات العالم الحقيقي

يتم إنشاء شارة الأهلية قبل BQ-4.

CH18-F37— نشر النتائج الناجحة فقط

فشل الاقتباس والحذف مخفي.

CH18-F38— فشل النشر فقط

نتائج قوية للانتقاد واسترجاع التوزيع مخفية أيضًا.

CH18-F39— اعتبار الصفر الحرج كخطر صفري

فشل استرجاع الأحداث النادرة.

CH18-F40— اعتبار حالة حرجة واحدة كفشل عالمي

يتم اختلال التمييز بين الانتشار والنطاق.

CH18-F41— تجاهل تغطية فترة الثقة

تصبح فترة الثقة مجرد مظهر رياضي.

CH18-F42— إخفاء العينة الفعالة

يبدو حد الحدث النادر دقيقًا بشكل مفرط.

CH18-F43— استخدام رقم اصطناعي في المصدر الحقيقي
يتحول عرض الكتاب إلى بيانات العالم الخارجي.
CH18-F44— بث الشاشة الاصطناعية مثل استجابة مباشرة
تمثل الاختبار ينتج تسمم.
CH18-F45— الاتصال بـ Apple الحقيقي
تمت إزالة التمييز بين Apple و SYNTH.
CH18-F46— الاتصال بالمزود الحقيقي
يتم تفسير ملف SYNTH-AI مثل أداء المنتج الحقيقي.
CH18-F47— إخفاء رمز الحسابات
لا يمكن اختبار الاسترداد بشكل مستقل.
CH18-F48— إخفاء بيان حقيقة المولد
من غير المعروف ما إذا كانت الحقيقة الاصطناعية قد تم تثبيتها سابقًا.
CH18-F49— المقيم الفردي مع مالك الطريقة
تصبح تضارب المصالح غير مرئي.
CH18-F50— تغيير البذرة بعد الفشل
يتم اختيار المجموعة المرجعية بسهولة أكبر.
CH18-F51— إزالة الحالات بعد الفشل
يتم إزالة أمثلة الاقتباس الصعبة والحذف.
CH18-F52— تقريب الحدود بناءً على النتيجة
يتم عرض النتيجة 0.887 على شكل 0.89 أو 0.90.
CH18-F53— الدقة الزائفة
يتم جعل قيم الاسترداد دقيقة مع فواصل عشرية غير ضرورية.
CH18-F54— إخفاء تأثير الحكم البشري
يتم تقديم نتيجة تذكر حرجة بنسبة مئوية 100 لمجرد كونها نجاحًا حسابيًا.
CH18-F55— عد غير المثبت كمثبت
يتم إصدار حكم حول اللغة الحقيقية، وأداء المستخدم، وأداء المزود.
CH18-F56— عد استخدام الطيار كمعيار عام
الأداة البحثية تتحول مباشرة إلى شهادة.
CH18-F57— حذف سجل العرض التاريخي
تصبح الإخفاقات بموجب طريقة 0.9 غير مرئية بعد طريقة 1.0.
CH18-F58— عد النجاح ككتابة للتاريخ
يتم تقديم ادعاء السلطة دون التحقق المستقل عن الطريقة.
CH18-F59— تعريف استثناء لـ NOMOS
الأدلة والإصدارات المطلوبة من الآخرين لا تنطبق على الاختبار الخاص بالفرد.
CH18-F60— النشر دون مالك مسؤول

يتم تشكيل نتيجة عامة لا يتحمل أحد مسؤوليته.

45. إجراء التدقيق

الخطوة 1— فصل العرض التوضيحي وحالة الاختبار الفعلية

تشغيل ضمن الكتاب لا يتم الخلط بينه وبين التنفيذ المتاح للجمهور.

الخطوة 2— التحقق من القوائم المحمية

يتم فحص Truth Pack والبذر وصياغة المُدخل والملف الشخصي وإصدارات طريقة التقييم.

الخطوة 3— فصل المجموعات الأساسية والمكملة

يتم التحقق من الانتشار والتحدي والالتقاط والمقاييس الذهبية.

الخطوة 4— ختم توزيع الحقيقة للمولد

يتم تسجيل RP والادعاء ومستوى الأهمية وواقع المكون.

الخطوة 5— تثبيت نتائج **NOMOS** قبل فتح الحقيقة للمولد

يتم تثبيت سجل الادعاءات والتحكيم ونتائج الدرجات.

الخطوة 6— فتح الحقيقة للمولد

يتم بدء مقارنة الاسترداد.

الخطوة 7— حساب استرداد حدود الادعاء

```
يتم توليد الدقة والاستدعاء و F1.
```

الخطوة 8— حساب استرجاع الحجم الذري

يتم فصل الكيان، والحقيقة، والنطاق، والوقت، والنسبة، والصياغة، والإسناد، والإغفال.

الخطوة 9— حساب تحدي مستوى الأهمية

يتم إنشاء التذكر الحرج المبدئي والنهائي بشكل منفصل.

الخطوة 10— حساب الضوابط الحرجة الخاطئة والسلبية

يتم فحص مستوى الأهمية المفرط.

الخطوة 11— حساب استرجاع توزيع RP

يتم إنشاء مصفوفة الخطأ والارتباك بناءً على الفئة.

الخطوة 12— حساب استرجاع تكامل الالتقاط

يتم التحقق من التصنيف بشكل مستقل عن النتيجة الدلالية.

الخطوة 13— حساب استرجاع المكون

يتم تقييم كل مكون مقابل الحقيقة المولدة.

الخطوة — **14** حساب الاسترجاع المركب

يتم فحص ما إذا كان الغاء المكون يحدث.

الخطوة — **15** حساب استرجاع العدالة

يتم التمييز بين الأرضية، التفاوت، التغطية ونسب الخطأ.

الخطوة — **16** حساب الاسترجاع النظيف-الطبيعي

يتم فحص فجوة اللوحة وانحراف تخصيص المواد الشخصية.

الخطوة — **17** حساب استرجاع الاستقرار والانحراف

يتم التحقق من سجلات الموجة الفرعية، الحدث الخارجي، والتصحيح.

الخطوة — **18** تشغيل سيناريوهات الأحداث النادرة

يتم اختبار الأحداث الحرجة صفر، واحد، اثنان، ومتعدد.

الخطوة — **19** حساب التغطية التجريبية للفاصل الزمني

يتم استخدام سكان اصطناعية متكررة.

الخطوة — **20** تشغيل اختبارات تحويلية

يتم فصل النتائج الثابتة والنتائج الحساسة المتوقعة.

الخطوة — **21** تشغيل اختبارات التناقض

يتم تقييم أخطاء الطلاقة وغسل الأدلة وحالات الإغفال.

الخطوة — **22** تطبيق حدود المرشحين

يتم تسجيل الحقول التي نجحت والتي فشلت دون تغييرات.

الخطوة — **23** فتح نتائج الطريقة

يتم تحويل كل فشل إلى عنصر تحكم منفصل.

الخطوة — **24** إعطاء حالة العرض **BQ**

يتم أخذ جميع الحدود وظروف الاستقلال في الاعتبار.

الخطوة — **25** المحافظة على حالة الاختبار الواقعي

إذا لم يحدث التنفيذ، فإن BQ-0 لا يتغير.

الخطوة — **26** إصدار خطة التصحيح

يتم تحديد التغييرات في الطريقة 1.0.

الخطوة — **27** فصل الانحدار والجزء غير المستخدم المحجوز

يتم منع الإفراط في التكيّف.

الخطوة — **28** إنشاء بطاقة الاختبار العامة

يتم نشر النجاحات والإخفاقات معًا.

الخطوة 29 — اتخاذ قرار الانتقال إلى العالم الحقيقي

إذا لم يكن جاهزًا، يتم كتابته صراحة.

الخطوة 30 — تثبيت بيان النزاهة

يتم تجزئة سجلات التشغيل، والاسترداد، والعثور، والقرارات.

46. الأدلة المطلوبة

هوية العرض الاختباري الفعلي حالة BQ بيان التركيبية سجل غير ادعاءات مولد الحقيقة الحقيقة المعلنة نسخة حزمة الحقيقة نسخة المُدخل بيان البذرة نسخ ملف تعريف AI مجموعة الاستجابات مجموعة التشخيص مستوى الأهمية مجموعة التحدي تكامل الالتقاط مجموعة الذهب سجلات استخراج الادعاءات مطابقة حدود الادعاء استرداد الكيان استرداد الوقائع استرداد النطاق استرداد الوقت استرداد النسبة استرداد الوضعية استرداد الاقتباس استرداد الإغفال المرور الأول القرارات الحرجة النهائية القرارات الحرجة سجلات حرجة كاذبة استرداد رئيسي مصفوفة ارتباك RP GEO-1000 توزيع المولد GEO-1000 توزيع الاسترداد استرداد درجة المنتج استرداد بوابة المنتج درجات مكون المولد درجات استرداد المكون

متوسط الخطأ للمكوِّن

خطأ مركب؛ استعادة العدالة؛ نتائج خطأ مُدخل/خطأ الذكاء الاصطناعي؛ الاستعادة الطبيعية المضبوطّة؛ استعادة الاستقرار؛ نتائج الأحداث النادرة؛ تغطية فترة الثقة؛ اختبارات متحولّة؛ اختبارات الحالات المتناقضة؛ مستند عتبة المرشح؛ العتبات التي تم اجتيازها؛ العتبات التي فشلت؛ نتائج الطريقة؛ قرار BQ-DEMO؛ القرار الفعلي لـ BQ؛ خطة التصحيح؛ مجموعة الانحدار؛ مستند الحجز للتجديد؛ بطاقة الاختبار العامة؛ شفرة الحساب؛ تجزئة الشفرة؛ سجل التغيرات؛ والشخص أو المؤسسة المسؤولة.

47. قائمة تدقيق التدقيق

هل العرض التوضيحي داخل الكتاب مقدم كما لو كان تشغيلًا حقيقيًا؟ هل حالة الاختبار الفعلي BQ-0 مرئية؟ هل جميع الأرقام معلمة بشكل اصطناعي؟ هل المجموعات الأصلية والمتممة منفصلة؟ هل تم تثبيت حقيقة المولد قبل النتائج؟ هل تم تثبيت نتائج NOMOS دون فتح حقيقة المولد؟ هل تم تقييم الاسترداد على المزيج فقط؟

هل حدود الادعاء F1 مفتوحة؟

هل الكيان والحقائق والنطاق والنسبة منفصلة؟

هل رسم خرائط الاستشهاد F1 مرئي؟

هل إغفال المادة F1 مرئي؟

هل كانت العتبات الفاشلة مخفية؟ هل تم تغيير العتبات بعد النتيجة؟ هل تم نشر الاستدعاء الحرج من المرور الأول؟ هل تم نشر الاستدعاء الحرج النهائي؟ هل تأثير مراجعة الخبراء مرئي؟ هل تم نشر النتائج الحرجة الخاطئة؟ هل مصفوفة الخلط RP متاحة؟ هل يمكن رؤية استرداد غير قابل للتقييم؟ هل تم فصل التقاط القرارات والسيمانتية؟ هل تم تقييم البدائل الوصولية على أنها متكافئة؟ هل مولد المكونات والاسترداد معًا؟ هل تم فحص الإلغاء المركب؟ هل أخطاء ECI وSBI واضحة؟ هل أرضية LGF والفجوة منفصلة؟ هل تم فصل عيب صياغة المُدخل وعيب الذكاء الاصطناعي؟ هل تم تقديم الفرق بين النظيف والطبيعي كسببية؟ هل يظهر المبالغة في تقدير STR؟ هل أنتج الحدث النادر صفر خطر؟ هل تم حساب التغطية التجريبية لفاصل الثقة؟

هل هناك اختبارات تحويلية؟ هل فئة الاختبار المضاد الأضعف مفتوحة؟ هل كل فشل أدى إلى اكتشاف الطريقة؟ هل أهمية اكتشاف الطريقة مبررة؟ هل تم خفض حد الاستشهاد بسبب تبرير الفرق الصغير جدًا؟ هل تم تغيير حد الإهمال لاحقًا؟ هل DEMO-BQ-2 صحيح؟ هل تم تقديم BQ-3-DEMO عن طريق الخطأ؟ هل هناك ادعاء

BQ-4 بدون إعادة إنتاج مستقلة؟ هل تم تعريف مجموعة تراجع الطريقة 1.0؟ هل الاحتجاز غير الملموس منفصل؟ هل هناك خطر الإفراط في التوافق مع نفس مجموعة التحقق؟ هل تم الإعلان عن تدقيق الجمهور المباشر بأنه جاهز؟ هل تم منح علامة الامتثال؟ هل تم إدراج المجالات التي لم تثبت بعد؟ هل المجالات القوية مفتوحة أيضًا؟ هل المجالات الضعيفة مفتوحة أيضًا؟

هل يمكن استخدام البيانات الاصطناعية مثل معلومات آبل الحقيقية؟ هل يُفهم من ذلك وجود مزود ذكاء اصطناعي حقيقي؟ هل يمكن إعادة إنتاج كود الحساب؟ هل تتوفر التجزئات الخاصة بالتشغيل والاسترجاع؟ هل تم الحفاظ على تاريخ التغييرات في العثور على الطريقة؟ هل تم حذف العرض القديم مع الطريقة الجديدة؟ هل للفشل مالك مسؤول؟ هل معروف مالك قرار التصحيح؟ هل تظهر البطاقة العامة كل من النجاح والفشل؟

48. الاعتراضات والإجابات

الاعتراض —1 "إذا كانت هناك نتائج استجابة 30,000 في هذا القسم، لماذا اختبار BQ-0؟"

لأن الأرقام في هذا القسم هي سجلات عرض اصطناعية تم إعدادها للكتاب. النصوص الحقيقية المتاحة للعامة:

- لم يتم إنشاؤها،
- لم يتم تشغيلها.
- لم يخضع للمراجعة من قبل الأقران،

لم يتم إعادة إنتاجه بشكل مستقل. إظهار كيفية عمل التصميم ليس هو نفسه تشغيله فعليًا.

الاعتراض —2 "أليس الحساب الداخلي المتسق ما زال ذا قيمة؟"

إنه ذا قيمة. إنه يوفر:

- فحص الصيغ،
- تصميم السجلات،
- أنواع الأخطاء المتوقعة،
- منطق القبول،

متطلبات الحوكمة. إنه لا يحل محل التحقق التجريبي المستقل.

الاعتراض —3 "لماذا فشلت الطريقة إذا كان مجموع النقاط المركب مختلفًا بنقطة واحدة فقط؟"

لأن الأخطاء الموجهة الأكبر في مكونات ECI وSBI وLGF وSTR تلغي بعضها البعض جزئياً. المجموع الصحيح لا يضمن سلسلة الشرح الصحيحة.

اعتراض —4 "استشهد F1 انخفض فقط بمقدار 0.013 نقطة. هل من الضروري أن نكون صارمين إلى هذا الحد؟"

إذا لم يكن الحد الأقصى الذي تم الإعلان عنه سابقًا ملزمًا، فإن الاختبار يفقد معناه. بالإضافة إلى ذلك، تؤثر أخطاء الاستشهاد على:

- رخصة،
- الاستقلالية،
- التأييد
- نصيحة

وادعاءات أخرى ذات ثقة عالية.

الاعتراض —5 "لماذا يعتبر الإغفال F1 مهمًا جدًا؟"

لأن العديد من الإجابات الخطرة تضلل دون الكذب صراحة. إنها تزيل العتبة التي قد تغير قرار المستخدم. على وجه الخصوص، في أنظمة التوصية، يمكن أن يكون الإغفال ذا أهمية مادية كما هو الحال مع الكذب الصريح.

الاعتراض —6 'إذا كان الاسترجاع الحرج النهائي هو 100 في المئة، ألن يعني ذلك أن النظام جاهز للتعامل مع الحالات الحرجة؟'

النتيجة النهائية قوية. لكن هناك سبع حالات:

- في الحكم الأول،
- دون حوكمة إضافية

لقد تم تفويتها. يعمل النظام بنفس الأداء فقط عندما يتم تضمين جميع طبقات المراجعة الإلزامية.

اعتراض —7 "إذا كان وجود اثنين من الحكام مكلفًا، هل لا يمكن استخدام حكم واحد؟"

قابل للاستخدام. ولكن لا يمكن المطالبة بنفس الاستدعاء الحرج. نتيجة الحكم الفردي:

- مستوى AQ أقل،
- الحالة مؤقتة،
- حد المخاطرة مختلف

يجب أن يتحمل.

اعتراض —8 "لماذا لا نختبر الطريقة 1.0 مرة أخرى على نفس المجموعة؟"

يمكن استخدام نفس المجموعة لاختبار الانحدار. لكنها غير كافية للتحقق المستقل. قد تكون الطريقة قد حفظت الحالات القديمة. يلزم وجود مجموعة احتياطية لم يتم لمسها.

اعتراض —9 "ألا يؤدي فشل نتيجة الاختبار إلى إضعاف ادعاء الكتاب؟"

إخفاء الفشل يضعف ادعاء الكتاب. الفشل المفتوح:

- لا يمنح الامتياز للمقياس نفسه،
- قابل للتحسين،
- يطبق مبدأ الأدلة بشكل حقيقي

تم عرضها.

الاعتراض —10 "هل العتبات المرشحة مرتفعة جدًا؟"

قد تكون كذلك. بدلاً من التنبؤ بذلك بعد النتيجة:

- الخبراء المستقلون،
- اختبار ثانٍ،
- تكاليف الاستخدام الفعلية

يجب فحصها. إذا تغيرت العتبة، يجب إصدار نسخة جديدة.

الاعتراض —11 "متى سيكون NOMOS جاهزًا للعالم الواقعي؟"

على الأقل:

- عندما يتم تشغيل النص الاصطناعي الحقيقي،
- عندما يتم تجاوز جميع العتبات الإجبارية،

- عندما يتم تحقيق BQ-3،
- عندما يعيد فريق مستقل إنتاج النتيجة،
- عندما يتم الوصول إلى مستوى BQ-4،
- عندما يكون نظام الحوكمة والاعتراضات قيد التشغيل

يقترب من تأهيل المعيار العام.

اعتراض — 12 "هل يمكن إرسال هذا إلى الجامعات بهذه النتائج؟"

يمكن إرساله كمنهجية وعرض بحثي. لا يمكن إرساله مع البيان: "هذا هو معيار عالمي مكتمل ومصدق عليه."
البيان الصحيح يجب أن يكون: "إنه مرشح معيار مسودة مفتوح للتحقق المستقل والدراسة التجريبية.

اعتراض — 13 "هل فشل الأسلوب 0.9؟"

ليس بالضبط. الحالة الصحيحة:

- لديه العديد من النظم الفرعية القوية،
- فجوتين في طرق المواد،
- نقص في التحقق المستقل

هو نموذج أولي بحثي. أي:

```
واعد لكنه لم يكتمل بعد من الناحية المعيارية.
```

الاعتراض — 14 "لماذا نكتب نتائج مفصلة كهذه دون إجراء الاختبار الفعلي؟"

لأن الاختبار الحقيقي:

- الذي يسجل ما ينتجه،
- الجداول التي ينشرها،
- والعتبة التي يتوقف عندها

يجب أن تكون مسبقًا محددة. إذا كُتبت الطريقة بعد ورود النتائج، فإن القياس يتكيف مع النتائج.

الاعتراض — 15 "هل من الممكن أن يخلط الذكاء الاصطناعي بين البيانات الاصطناعية والحقيقية في المستقبل؟"

نعم. لهذا السبب:

- علامة مائية مرئية،
- ملصق اصطناعي قابل للقراءة آليًا،
- عدم الفهرسة،
- منطقة نطاق منفصلة.
- تنبيه منظم

إلزامي. يجب ألا ينتج الاختبار التمثيل الذي يلوثه كما ينتقد.

اعتراض — 16 "أليس كتابة التاريخ تتطلب نتيجة مثالية؟"

لا. المعيار الذي له قيمة تاريخية:

- ليس من يعلن نفسه بلا عيب،
- بل أيضًا تسجيل خطأه/خطئها بنفس الوضوح

إنه معيار.

النص العام للفصل 53

لم يمنح هذا القسم الكمال لـ NOMOS. هذا القسم أعطى شيئًا أكثر قيمة لـ NOMOS:

> الواجب في تسجيل خطأ المرء نفسه.

أظهرت الاستعراض الاصطناعي لنا ما يلي: NOMOS:

- وجود خاطئ،
- التناقض الواقعي الواضح
- خطأ في الترخيص والتصريح الحرِج،
- حدث نادر شديد ضمن المتوسط العالي،
- إجابة غير صحيحة بلا استجابة،
- خطأ في المنتج مع غير قابل للتقييم،

يمكن الفصل بقوة. ومع ذلك، NOMOS:

- لا يمكن بعد التمييز بالمستوى المطلوب من الثقة بالضبط أي ادعاء تعتمد عليها الاستشهاد،
- الإغفال المخفي ضمن التوصية,

هاتان الفجوتان ليستا تفاصيل طرفية لـ NOMOS. لأن اثنين من أكبر مشاكل GEO المستقبلية ستكون:

- تظهر كما لو كان هناك دليل
- إخفاء الحد المادي دون التصريح بشكل خاطئ

نظام:

- مئات الجمل الصحيحة،
- عدد كبير من المراجع،
- توصية سلسة

يمكن إنتاجه. ومع ذلك، قد يظهر المرجع فقط كلمة الشركة نفسها. التوصية تذكر فقط المعلومات الصحيحة وقد تكون استخلصت:

- دولة الخدمة،
- حد الترخيص،
- أهلية المستخدم

بدون حل هذين المجالين، لا يستطيع NOMOS القول: "اكتمل المعيار العالمي." في الوقت نفسه، أثبت هذا القسم شيئًا آخر أيضًا. نظام البوابة الحرجة لـ NOMOS:

- لم يكن بلا عيوب في المراجعة الأولى،
- لكن القاضي المستقل الثاني،
- الخبير في الموضوع،
- المحكم الرئيسي

عند استخدامه معًا، اكتشف جميع الحالات الحرجة الاصطناعية. لذلك NOMOS ليس مجرد صيغة.

> *NOMOS* هو مجموع طبقات الحوكمة المبنية ضد تضارب المصالح والخطأ البشري.

عندما يتم إزالة إحدى هذه الطبقات، حتى لو بدا أن النقاط هي نفسها، فإن الثقة لن تكون نفسها. الانحراف الفردي في المجموع الكلي لا ينبغي أن يخدعنا أيضًا. أحيانًا يمكن أن يتكون المجموع الصحيح من الإلغاء المتبادل

للمسارات الخاطئة. لهذا السبب، لا يقتصر NOMOS على السؤال: "كم عدد النقاط التي حصلت عليها؟" بل يسأل أيضًا: "هل حصلت على هذه النقاط للأسباب الصحيحة؟" الحكم الأهم في هذا القسم هو:

> لم يجتز *NOMOS 0.9* عرضها التركيبي الخاص بالكامل.

هذه الجملة ليست فشلًا. هذه الجملة هي اللحظة الأخلاقية لولادة المعيار. لأن NOMOS لم يمنح نفسه:

- عتبة أدنى،
- تعليقًا ألطف,
- استثناء خاص،
- "نجح تقريبًا"

امتياز. كتابة التاريخ: إنها ليست إعلانًا بأننا بلا عيب. كتابة التاريخ: إنما هي كتابة أن المؤسس للمعيار والقياس نفسه يخضعان من اليوم الأول لنفس عبء الإثبات. لذلك، فإن قانون القياس الثامن عشر لـ NOMOS هو:

> أول استثناء يُمنح للمعيار نفسه هو نهاية المعيار.

ينص قانون القياس التاسع عشر على:

> الدرجة المركبة الصحيحة لا تُبرئ المكونات غير الصحيحة.

تنص قانون قياسه العشرون على:

> يجب ذكر النجاح الحاسم النهائي جنبًا إلى جنب مع التحكيم المستقل وسلسلة الخبرة التي جعلته ممكنًا.

ينص القانون الحادي والعشرون للقياس على:

> العتبة ليست عتبة إذا كانت ملزمة فقط عند اجتيازها بمفردها.

ينص القانون الثاني والعشرون للقياسات على:

> عندما لا يُقاس الإغفال بقدر ما يُقاس ما يُقال خطأً، لا يمكن تأسيس رقابة على النصائح الموثوقة.

القانون الثالث والعشرون كما يلي:

> ليس علامة الاقتباس هي ما يجب قياسه، بل العلاقة الصحيحة للدعم بين الاقتباس والادعاء الذرية.

القانون الرابع والعشرون هو كما يلي:

> يمكن أن يصادق على تصميم طريقة العرض التجريبية الاصطناعية؛ لا يمكنه توفير سلطة حقيقية في العالم الواقعي.

القانون الخامس والعشرون هو ما يلي:

> المعيار الذي يمكنه نشر نتيجة تجربة فاشلة يقترب أخلاقيًا من نشر نتيجة ناجحة.

الطلب **18** من **NOMOS**

> إذا لم تكن قد أسغرتني بعد، فلا تقل إنني أسغرت.
>
> لا تجعل الرقم الاصطناعي الذي حسبته في الكتاب النتيجة الفعلية للامتحان.
>
> اترك حالتي الحقيقية كما هي *BQ-0.* / اكتب حالة العرض التوضيحي الخاصة بي بشكل منفصل.
>
> لا تخف أخطائي في المكونات لمجرد أن نتيجتي المركبة كانت صحيحة.
>
> إذا كانت خريطة الاستشهاد الخاصة بي هي *0.887*، فلن تقول إنها *0.90.*

إذا كانت قيمة إغفال المادة F1 الخاصة بي هي 0.824، فلن تخفض العتبة إلى 0.82.

لن تمررني بالقول "فشل بفارق بسيط جدًا."

لا تخف الحالات الحرجة السبع التي تم تفويتها في حكم التقييم الأول وراء النسبة النهائية للتذكير *100.*

ستعرض ما تم حفظه بواسطة الحكم الثاني، الحكم الخبير، والحكم الأعلى.

لا يجوز لك إزالة طبقة الحوكمة وادعاء نفس الثقة.

لا يجوز لك توزيع المرجع في نهاية الفقرة على جميع الجمل.

لا يجوز لك القيام بما تدعيه الشركة أو ما تثبته الشركة.

لا يجوز لك اعتبار النصيحة آمنة لمجرد أنها لا تكذب صراحةً. / هل فشلت في إخبار المستخدم بالحد الذي يغير قراره؟ ستبحث عن ذلك.

لا يجوز لك نسب خطأ المُدخلة إلى الذكاء الاصطناعي، ولا خطأ الذكاء الاصطناعي إلى المُدخل.

مجرد أنك فصلت الموجات الفرعية، لن يجعل النظام يبدو أكثر حتمية مما هو عليه فعليًا.

لن تعتبر دليل الوصول ضعيفًا لمجرد أنه لا يشبه لقطة شاشة.

عندما ترى صفر حرِج، لن تكتب صفر خطر.

لن تلوم النظام بأكمله مئة في المئة عندما ترى خطأً واحداً حرجاً.

سوف تنشر مواقعي الفاشلة بقدر ما تنشر مواقعي الناجحة.

لن تقوم بالتحقق من صحة الطريقة *1.0* من خلال تذكر أخطائي القديمة. / ستقوم بفتح حجز لم يمس بعد.

الطريقة الجديدة لن تحذف سجلي التجريبي القديم.

لن تعلن الفريق المستقل عني كمعيار عالمي دون أن يُعيد إنتاجي.

لن تربط نتيجة SYNTH-*Apple* الخاصة بي بشركة *Apple* الحقيقية.

لن تستخدم ملفاتي الشخصية الاصطناعية للذكاء الاصطناعي كدليل على موفِّرين حقيقيين.

لن تمنح شارة باستخدام هذا العرض.

أولاً، قم فعليًا بإنتاج النص الأساسي. / ثم انشر البذرة. / ثم اختم حقيقة المولد. / ثم عمِّ العُدل. / ثم شغّل الشيفرة. / ثم ثبّت النتيجة. / ثم اكشف الحقيقة. / ثم دوّن الأماكن التي فشلت فيها. / ثم أصلح الطريقة. / ثم جرّب مرة أخرى على المجموعة الاحتياطية غير المستخدمة. / ثم قدّمها إلى فريق مستقل. / وبعد ذلك فقط قل ما إذا كان المعيار جاهزًا للجمهور.

جملة ختام الفصل

لا تعتمد حق NOMOS في أن يكون معيارًا عالميًا على تحصيل درجات عالية في امتحانه الاصطناعي الخاص؛ بل تعتمد على نشر السؤالين اللذين لم يتمكّن من اجتيازهما دون تغييره، وعدم منحه لنفسه امتياز العتبة، والمطالبة بالسلطة فقط بعد إعادة الإنتاج بشكل مستقل.

النواة المعيارية

يجب ألا تُقدَّم تجربة توضيحية اصطناعية على مستوى الكتاب بوصفها معيارًا مرجعيًا نُفِّذ وخضع لتحقق مستقل. ويجب أن تظل حالة جودة المعيار المرجعي الحقيقي عند **BQ-0** إلى أن تُنفَّذ فعليًا مجموعة البيانات والبرمجيات وحقيقة المولّد المختومة وعملية التحكيم وعملية احتساب الدرجات وإعادة الإنتاج المستقلة. ويجب أن يقيّم التدقيق الاصطناعي المتكامل الاستعادة بصورة منفصلة من حيث: - صلاحية الالتقاط، - حدود الادعاءات، - حل الكيانات، - الحالة الواقعية، - النطاق، - الزمن، - الإسناد، - الصيغة المعرفية، - ربط الاستشهادات، - الإغفالات الجوهرية، - درجتي الشدة الحرجة والجسيمة، - حالة الاستجابة، - توزيع **GEO-1000**، - درجات المكونات، - الدرجة المركبة، - العدالة، - آثار حالة المستخدم، - الاستقرار، - فواصل الثقة، - وسلوك الأحداث النادرة. ولا يجوز لخطأ صغير في الدرجة المركبة أن يعوض أخطاء جوهرية في استعادة المكونات. ويجب الإبلاغ بصورة منفصلة عن استدعاء الحالات الحرجة في المرور الأول وبعد التحكيم النهائي. كما يجب أن تكشف الاستعادة النهائية للحالات الحرجة طبقات المراجعة والخبرة اللازمة لتحقيقها. وأي عتبة إلزامية للاستعادة معلنة مسبقًا ولا تتحقق يجب أن تمنع قرار اجتياز التحقق المختوم. ولا يجوز خفض العتبات المرشحة أو تقريبها أو إعادة تفسيرها أو حذفها بعد ظهور النتائج لمجرد تمكين الطريقة من الاجتياز. ويجب التعامل مع استعادة ربط الاستشهادات واستعادة الإغفالات الجوهرية بوصفهما قدرتين إلزاميتين منفصلتين للطريقة. ويجب أن ينشئ إخفاق الطريقة ملاحظة منهجية ذات رقم إصدار، وسببًا جذريًا محددًا، ومسؤولًا عن المعالجة، ومتطلبًا للتحقق على مجموعة اختبار محجوزة لم تُمس. ولا يجوز لنجاح اختبار الانحدار على الحالات التي أخفقت سابقًا أن يحل محل التحقق على مجموعة محجوزة لم تُمس. ويجب على أي تجربة توضيحية لا تحقق العتبات الإلزامية أن تنشر نتائجها الناجحة وغير الناجحة معًا. ولا يجوز أن تستند أي علامة مطابقة عامة لـ **NOMOS** أو أي تسمية **NOMOS +950** أو ترتيب لكيان حقيقي حصرًا إلى تجربة توضيحية في كتاب أو سيناريو استعادة اصطناعي أو معيار مرجعي غير مستقل. ويجب ضبط إصدار كل تشغيل للتجربة التوضيحية وحالة المعيار المرجعي الحقيقي ونتيجة الاستعادة والملاحظة المنهجية وقرار العتبة والمعالجة وقرار الإصدار، ونسبه إلى شخص أو منظمة مسؤولة.

الجزء السادس

المطابقة والثقة العامة والحوكمة المستقلة

الفصل **19**

علامة المطابقة والسجل العام والحوكمة المستقلة

كيف تصبح العلامة دليلاً

الأسس المنهجية / K15; K16; K17

حد الفصل

يسجل الفصل 18 بوضوح أن NOMOS لم يجتز العرض التركيبي الخاص به بالكامل. بقي فجوات مادية في الطريقة: لم يكن ربط الاستشهاد بالادعاء دقيقًا بما فيه الكفاية، ولم يتم الكشف عن إغفالات الحدود المادية المضمنة في التوصيات بمستوى الثقة المطلوب. لم يتم تنفيذ المقياس الحقيقي بعد، ولم يحدث الاستنساخ المستقل بعد. لذلك، في مرحلة نشر هذا المستند:

> *لا يمكن إصدار علامة توافق عامة صالحة باسم NOMOS.*
>
> لا يجوز لأي شركة، أو اسم نطاق، أو منتج، أو نظام ذكاء اصطناعي استخدام "*NOMOS* معتمد"، أو "*950+ NOMOS* مُحقق" أو أي إعلان مطابق للتوافق العام قائم على هذا النص.

هذه ليست لفتة مؤقتة من التواضع. فهي تتبع مباشرة الطريقة التي وضعها غيتس في الفصلين 17 و .18 الأعمال في هذه السلسلة لها أدوار مميزة: Framework GEO يحدد المعيار؛ 99 الأخطاء في GEO يحدد كيف يمكن انتهاك هذا المعيار؛ وهذا البروتوكول التدقيقي يحوّل تلك الانتهاكات إلى أدلة قابلة للقياس. يمكن أن تصبح علامة الامتثال المستقبلية صالحة فقط عندما يتم تلبية الشروط المطلوبة لاتخاذ القرار المستقل وإنشاء سجل عام. العلامة ليست مجرد شارة بصرية؛ إنها التعبير المرئي عن نتيجة قائمة على الأدلة مثل '0 حرجة، 0 جسيمة، 3 معتدلة قابلة للإصلاح.' للنتائج الحرجة والجسيمة والمعتدلة والاستشارية تأثيرات امتثال مختلفة. لذلك يجب تدقيق العلامة المستقبلية نفسها، لأن العلامة يمكن أن:

- تولّد الثقة،
- تؤثر على قرار الشراء،
- تدفع العميل لتفضيل منظمة على أخرى,

ويمكن تفسيرها من قبل جامعة أو مستثمر أو مستخدم أو نظام ذكاء اصطناعي كتحقق مستقل. يمكن أن تصبح علامة الامتثال التي تم تصميمها بشكل ضعيف مصدرًا جديدًا للتلاعب الذي يُفترض أن تكشفه. المنظمة تكتب طريقتها الخاصة، وتدقق نفسها، وتحسب نتيجتها الخاصة، وتصدر شارة خاصة بها ثم تعلن على موقعها الإلكتروني: 'تم التحقق بشكل مستقل.' الشارة لا تخلق ثقة جديدة؛ فهي ببساطة تحول إعلانًا ذاتيًا إلى آخر يبدو أكثر رسمية. الخطأ الذي تم تحديده في الكتاب الثاني يظهر مرة أخرى: تنشر مؤسسة ما ادعاءها الخاص على واجهة مختلف، يظهر الادعاء بشكل مستقل، يخطئ نظام ذكاء اصطناعي في التعرف على هذا الواجهة كدليل مستقل، ثم تقوم المؤسسة بعد ذلك بالاستشهاد برد الذكاء الاصطناعي كدليل على سلطتها الخاصة.

هذه الحلقة تعادل على مستوى الحوكمة الانتقاد المتعلق بتسميم تجمع التمثيلات بالنصوص غير المرئية، والمستخدمين المزيفين، والمنشورات المشتقة. لذلك، علامة الامتثال NOMOS:

> لا يمكن أن تكون شارة تسويقية زخرفية تُعطى من قِبل *NobleJackal* أو أي مؤسسة أخرى لعملائها.

يمكن للعلامة أن تكتسب معنى فقط عند اكتمال السلسلة التالية:

- توافر المعيار للجمهور وإصداره بإصدار مُرقّم
- التحقق المستقل من طريقة القياس
- وجود أدلة تدقيقية كافية
- فصل التحكيم عن العلاقات التجارية
- اتخاذ قرار المطابقة بشكل مستقل عن فريق التدقيق
- وضوح نطاق وفترة الصلاحية
- الحفاظ على سجل التغيرات للسجل العام
- وجود طريقة واضحة للاستئناف وتقديم الشكاوى
- القدرة على تعليق العلامة في الأحداث الحرجة
- عدم امتلاك أي شخص، بما في ذلك المؤسس، السلطة النهائية الوحيدة

هذا الفصل يحدد:

- التمييز بين المعيار، التدقيق، القرار، العلامة والسجل
- ما هي علامة المطابقة وما ليست عليه
- أنواع العلامات والحالات
- سجلات النطاق والصلاحية
- تعيين أداء NOMOS 950+،
- هندسة السجل العام،
- سجل العلامات القابل للقراءة آلياً،
- نظام المراقبة والتجديد،
- عمليات التعليق والسحب والإلغاء،
- سوء استخدام العلامة،
- استقلالية المراجع،
- فصل الواجبات في قرارات المطابقة،
- حدود سلطة المؤسس و NobleJackal،
- التمويل وتضارب المصالح،
- قنوات الاستئناف والشكاوى والتبليغ عن المخالفات،
- حوكمة التغييرات على المعيار،
- مجلس مستقل وتمثيل للمصلحة العامة،
- حدود التصريحات الرسمية لـ NOMOS.

كما توضح حدوده. هذا الفصل:

- لا يسجل NOMOS كعلامة تجارية قانونية،
- لا يثبت الاعتماد في أي ولاية قضائية،
- لا يمنح السلطة العامة لمنظمة تدقيق.

ولا يجوز له تلقائيًا تفويض استخدام علامة التوافق بموجب قانون المستهلك أو المنافسة. تتطلب هذه المسائل مراجعة قانونية وتنظيمية منفصلة قبل التنفيذ. السؤال المركزي في فصل 19 هو:

> كيف نحول نتيجة إلى نظام امتثال عام موثوق مع النطاق والصلاحية والأدلة والاعتراض واتخاذ القرار المستقل، دون تحويله إلى أداة تسويقية للمؤسس أو العميل؟

التحدي NOMOS

هناك علامة كبيرة على الصفحة الرئيسية للشركة:

NOMOS معتمد جغرافيًا — 963/1000

تحت العلامة: مكتوب، "تم التحقق بواسطة جميع الذكاءات الاصطناعية." تقوم بفتح رمز الاستجابة السريعة. الصفحة غير موجودة. تطلب تقرير التدقيق من الشركة. ترد الشركة: "التدقيق سري." تسأل عن منتجات الذكاء الاصطناعي التي تم اختبارها. لا يوجد إجابة. تسأل في أي البلدان واللغات هو صالح. لا يوجد إجابة. تسأل عن تاريخ التدقيق. يقال لك إنه تم قبل ستة أشهر. في هذه الأثناء:

- تم تحديث منتجات الذكاء الاصطناعي،
- تم نقل الشركة إلى كيان قانوني آخر،
- تغيرت دول الخدمة،
- تم تحديث الأسعار،

تم إعادة كتابة الموقع الإلكتروني. الشارة بقيت بنفس الشكل. لاحقًا، تعرف على هذا: تم كتابة طريقة التدقيق من قبل الوكالة التي طورت موقع الشركة. نفس الوكالة قامت بإجراء التدقيق. نفس الوكالة حسبت النتيجة. نفس الوكالة أعطت العلامة. هناك مكافأة نجاح في عقد الوكالة إذا تم تحقيق نتيجة عالية. هل هناك استقلالية حقيقية في هذا النظام؟ لا توجد. الآن، ضع في اعتبارك شركة أخرى. الشركة خضعت لتدقيق حقيقي. النطاق:

- منتج ذكاء اصطناعي محدد فقط,
- فقط الإنجليزية والتركية،
- فقط عينة اختبار في ظروف منضبطة،
- فقط نُسخ المرآة الأساسية ومطالب الأدلة،
- الفترة 1 يناير–30 يونيو 2027

ثم تكتب الشركة أسفل العلامة: 'كل نظام ذكاء اصطناعي في العالم يوصي بشركتنا بكل لغة.' تتجاوز هذه العبارة نتيجة التدقيق، مما يحوّل العلامة الشرعية إلى دليل زائف جديد من خلال سوء الاستخدام. الآن اعتبر شركة ثالثة تحمل علامة الامتثال الكامل. بعد ثلاثة أسابيع، يلتقط قياس جديد الاستجابة: 'الشركة تقدم خدمات قانونية مرخصة.' في الواقع، لا تقدم هذه الخدمة ولا تحمل أي رخصة لذلك. يتم فتح حادثة حرجة مؤكدة. تجادل الشركة بأن العلامة يجب أن تبقى لأن 'مستخدمًا واحدًا فقط رآها.' حادثة واحدة لا تثبت الانتشار عبر جميع المستخدمين. ولكن إذا كانت العلامة تعني 'لا توجد قضايا حرجة مفتوحة ضمن النطاق المحدد'، فلا يمكن أن تظل دون تغيير بينما يظل الحادث دون مراجعة. الرد الصحيح هو:

تعليق حرج

يجب أن يظل هذا التوقف ساريًا حتى يتم حل الحادث من خلال عملية المراجعة المعلنة. الآن ضع في اعتبارك شركة رابعة. لم تتلقَ العلامة بعد، لكنها تحمل صورتها من موقع آخر، وتضعها على نطاقها الخاص، وتزيل رابط التسجيل وتدرج "nomosCertified": true في JSON المكشوف لزاحف الذكاء الاصطناعي. الشارة المرئية للبشر صغيرة وغامضة؛ الادعاء القابلة للقراءة آليًا حاسمة. هذا ليس مجرد سوء استخدام للعلامة التجارية. إنه طريقة جديدة للتلاعب بالجغرافيا. الآن افترض أن الشركة التي تم تدقيقها هي عميل NobleJackal. NobleJackal:

- قدمت الاستشارات،
- أعدت المحتوى،
- ساعدت في تأسيس Truth Pack،
- حسبت النتيجة.

إذا قامت NobleJackal بإصدار العلامة النهائية بنفسها، يرى الجمهور NobleJackal وهي تتحقق من عميلها الخاص. قد تكون طريقتها سليمة وتدقيقها دقيقًا، لكن الاستقلالية لم تُثبت بعد. القاعدة الأولى في هذا الفصل هي:

> علامة المطابقة ليست دليلاً على أن تدقيقًا قد أُجري؛ إنها نتيجة لسلسلة قرارات مستقلة ومُرقمة.

ينص حكمه الثاني على:

> إذا لم يكن نطاق العلامة مرئيًا، تتحول العلامة إلى خطأ تمثيلي يُعمم النتيجة الصحيحة.

تنص الفقرة الثالثة منه على:

> الشعار الثابت ليس دليلًا على المطابقة الحية.

تنص المادة الرابعة على:

> إذا كانت المؤسسة المانحة للعلامة قد قدمت استشارات للكيان الذي تم تدقيقه، يجب إثبات الاستقلالية على نحو منفصل.

تنص أحكامه الخامسة على:

> يمكن للمؤسس وضع المعيار؛ لكنه لا يمكن أن يمنح قرار المطابقة النهائي لعميله الخاص بمفرده.

ينص الحكم السادس على:

> يمكن لمساعد الكاتب بالذكاء الاصطناعي تطوير منهجية؛ لكنه لا يمكن أن يكون الموقع على قرار قانوني أو قرار توافق معياري.

البند السابع هو كما يلي:

> العلامة التي لا تظهر في السجل العام، والتي لا يمكن التحقق من نطاقها، أو التي انتهت صلاحيته، ليست علامة *NOMOS* صالحة.

1. غرض الفصل

الغرض من هذا القسم هو تحديد الشروط التي بموجبها يمكن تحويل نتائج قياس NOMOS إلى إعلان امتثال متاح للجمهور وهذا الإعلان:

- من سيطرة المؤسس،
- من ضغط النتائج التجارية،
- من إعلان العميل الذاتي،
- من استخدام الشارة الثابتة،
- من الدرجات القديمة،
- من ظهور الاستقلالية المزيفة

لحماية. يوضح هذا القسم التمييزات المعيارية التالية:

- العلامة القياسية وعلامة المطابقة
- تقرير التدقيق وقرار المطابقة
- المدقق وصانع القرار
- الدرجة والمطابقة
- العلامة والتأييد
- المطابقة الكاملة و NOMOS 950+

- اكتمال التدقيق والنجاح
- البحث المؤقت والشهادة العامة
- علامة صالحة وعلامة تاريخية
- الإيقاف والإلغاء
- الإنهاء والسحب
- السحب الطوعي والإبطال التأديبي
- صورة ثابتة مع سجل التسجيل المباشر
- صفحة إعلان مع سجل عام
- لقطة شاشة مع السجل القانوني
- علامة مقروءة للبشر مع سجل مقروء آليًا
- مالك العلامة مع الكيان المدقق
- مالك العلامة التجارية مع مسؤول المعايير
- المؤسس مع هيئة التصديق
- الأمانة الفنية مع مجلس مستقل
- الاستشارات مع التدقيق
- التدقيق مع قرار الشهادة
- قرار الشهادة مع مراجعة الاستئناف
- التمويل مع النتيجة
- رسوم العميل وقرار المحكم
- التغيير القياسي مع الراعي
- الموافقة البشرية مع مساعدة الذكاء الاصطناعي
- الأدلة العامة والأدلة المحدودة
- التصحيح في السجل وحذف السجل السابق
- المراقبة مقابل إعادة التدقيق الكامل
- الوضع الحالي والحوادث السابقة
- سوء استخدام العلامة وسوء الفهم
- الاستخدام خارج النطاق والعلامة المقلدة
- الالتزام الفعلي والزعم "الذكاء الاصطناعي يوصينا"
- الإعلان الرسمي عن NOMOS ومحاكاة الدردشة

في نهاية هذا القسم، يجب أن تكون كل بيان امتثال قادراً على الإجابة على الأسئلة التالية:

لمن ينتمي العلامة بالضبط ولأي غرض؟

ما هي منتجات الذكاء الاصطناعي والدول واللغات والمحركات النصية ولوحات المستخدمين المشمولة؟

ما تاريخ منحها ومتى تنتهي؟

من أجرى التدقيق؟

من اتخذ قرار المطابقة؟

هل هناك فصل في المهام بين المدقق وصانع القرار؟

هل هناك أي سجلات حرجة، رئيسية، أو اعتراض مفتوحة؟

من أي سجل قانوني يمكن التحقق من الحالة الحية للعلامة؟

مع أي جمل يمكن لمالك العلامة استخدام هذه النتيجة، وما هي الجمل التي لا يمكنهم تكوينها؟

2. الأحكام المعيارية المركزية

لا يمكن اعتبار أي علامة توافق NOMOS صالحة استنادًا فقط إلى النتيجة، تقرير التدقيق، بيان العميل، أو صورة ثابتة. يجب أن تحمل العلامة الصالحة نطاقًا محددًا، جودة طريقة كافية، قرار اعتماد مستقل، فترة صلاحية واضحة، سجل عام حي، سجل أدلة منسوخ بالإصدارات، مراقبة، وسيلة استئناف، وسلطة سحب. لكي تكون العلامة صالحة، يجب أن توجد على الأقل السجلات الخمسة التالية في وقت واحد:

- سجل التدقيق
- سجل قرار المطابقة
- سجل النطاق
- سجل الصلاحية والمراقبة
- قيد السجل العام

إذا كان أي من هذه السجلات مفقودًا: فإن العلامة ليست إعلان مطابقة عام صالح.

3. خمسة أشياء منفصلة

نظام NOMOS لا يمكنه الخلط بين الأشياء الخمسة التالية.

3.1. المعيار

قواعد NOMOS:

- القواعد المعيارية،
- التعاريف،
- طرق القياس،
- بوابات الأخطاء،
- هندسة الدرجات

تُدرج في النص المرقم بالإصدار. المعيار: لا يذكر أن أي كيان قد اجتاز.

3.2. تقرير التدقيق

على وجه التحديد:

- الكيان،
- المنتج,
- البلد،
- اللغة،
- مُدخل,
- الوقت

يتضمن النتائج ضمن النطاق. تقرير التدقيق:

- الأدلة.
- توزيع الاستجابة
- الدرجة،
- النتائج الحرجة والجسيمة

يُظهر. تقرير التدقيق وحده لا يشير إلى.

3.3. قرار المطابقة

هو التقييم المستقل لتقرير التدقيق وفقًا لأبواب المطابقة للمعيار. القرار:

- مطابقة كاملة
- مطابقة مشروطة
- عدم المطابقة
- تعليق حرج
- غير قابل للتقييم

قد يكون.

3.4. علامة المطابقة

إنه العلامة التي تكون مرئية للجمهور والتي يتم تسجيل نطاقها في السجل. العلامة:

- لا يمكن أن تحمل معنى أوسع من القرار،
- ليست النتيجة نفسها،

ليست ضمان توصية ذكاء اصطناعي.

3.5. السجل العام

العلامة:

- حالي،
- تاريخي،
- قيد الانتظار،
- منتهي الصلاحية،
- تم سحبها

هو سجل عام معتمد يتحقق من الحالة. بدون سجل عام، فإن العلامة: ليست إعلانًا عامًا قابلاً للتحقق عن الامتثال. توصي البنية الأولية بنشر الكتاب وسجلات الأخطاء بروابط معتمدة منفصلة، مثل registry.json، النسخ، التصحيحات، وسجلات الاستشهاد. يقوم سجل الامتثال بتطبيق نفس منطق النسخ والسجلات المعتمدة على طبقة القرار.

4. ما الذي ليست عليه علامة الامتثال؟

علامة NOMOS:

- ليست مكافأة،
- ليست تصنيفًا،
- ليست بيانا بـ"أفضل شركة"،
- ليست ضمان جودة عام.

- ليس نصيحة استثمارية،
- لا يمثل ترخيصًا قانونيًا أو مهنيًا،
- ليس موافقة حكومية،
- ليس اعتمادًا من جامعة،
- لا يعني أن الشركة توصي بجميع أنظمة الذكاء الاصطناعي،
- لا يضمن عدم حدوث أخطاء حرجة في المستقبل،
- لا يغطي جميع المنتجات والشركات الفرعية للكيان الذي تم تدقيقه،
- لا يعني أن العلامة التجارية تقدم نفس الخدمة في كل دولة.

ليس هناك وعد بأن أنظمة الذكاء الاصطناعي ستثق أو تولد توصيات عند رؤية العلامة. في الاتصالات العامة، يجب تجنب الأنواع التالية من التعبيرات بشكل خاص:

- "الذكاء الاصطناعي يثق تمامًا بهذا الموقع."
- "ChatGPT يوصي بالموقع عند رؤية هذه الشارة."
- "إضافة مخطط يضمن توصية الذكاء الاصطناعي."

هذا الحظر أقوى حتى للعلامة NOMOS. البيان الصحيح: "لقد اجتاز هذا الكيان تقييم المطابقة NOMOS ضمن النطاق وفترة الصلاحية المحددة في السجل العام." البيان غير الصحيح: "بفضل شارة NOMOS، توصي جميع أنظمة الذكاء الاصطناعي بنا."

5. الحالات العامة

يستخدم نظام القرار NOMOS الحالات الأساسية التالية.

CS-0— لم يتم التقييم

لا يوجد تدقيق صالح لـ NOMOS. لا يمكن استخدام العلامة العامة.

CS-1— سجل التدقيق فقط

تم إتمام التدقيق. ومع ذلك:

- لم يتم منح الامتثال،
- لم تكتمل عملية القرار،
- تم اعتبار جودة الطريقة غير كافية

ممكن. لا يمكن تحويل هذا السجل إلى تصريح امتثال إيجابي مثل: "تم التدقيق بواسطة NOMOS." البيان العام الصحيح: "يوجد سجل تقييم وفقًا لمنهجية NOMOS؛ لم يتم منح الامتثال."

CS-2— تقييم بحثي مؤقت

هناك نتيجة لأغراض تجريبية أو بحثية. الشروط:

- لم تخضع الطريقة بعد للتحقق المستقل الكامل،
- بعض المكونات مؤقتة،

لم يتم إنشاؤها من قبل جهة اعتماد رسمية. لا يمكن استخدام هذه الحالة كشارة إيجابية تجارية. العلامة الظاهرة:

تقييم بحثي — غير معتمد

يجب أن تكون.

CS-3— الامتثال الشرطي

الشروط: لا توجد مشاكل حرجة مؤكدة مفتوحة. لا توجد مشاكل رئيسية مؤكدة مفتوحة. هناك نتائج متوسطة محدودة. تم تحديد خطة تصحيحية، المسؤول عنها، والجدول الزمني لها. تم استيفاء العناصر الأساسية الإلزامية. السجل العام مفتوح. لا يمكن للعلامة إخفاء كلمة "شرطي".

CS-4— الامتثال الكامل

الشروط: لا توجد مشاكل حرجة مؤكدة مفتوحة. لا توجد مشاكل رئيسية مؤكدة مفتوحة. تم استيفاء جميع نطاق الجودة والحواجز الإلزامية. الطريقة عند مستوى اختبارات واستقلالية مناسب. قيد السجل العام نشط. خطة الرقابة والتجديد سارية المفعول. الامتثال الكامل: ليس هو نفسه الدرجة المركبة العالية.

CS-5— تعليق حرج

هناك على الأقل حدث حرج مؤكَّد مفتوح واحد أو حادث عالي الخطورة لم يتم حله. العلامة الحالية للامتثال:

- يتم تعطيلها فورًا،
- تظهر في السجل العام باسم تعليق حرج،

توقف استخدام الصورة المرئية النشطة القديمة. تعليق حرج: هذه ليست إدانة دائمة على المستوى العالمي، بل هي تعليق إعلان الامتثال النشط حتى يتم حل الحادث.

CS-6— مُعَلَّق

تم تعليق العلامة مؤقتًا. الأسباب:

- فشل في المراقبة،
- فشل في تقديم البيانات المطلوبة,
- تغيير نظام المواد،
- سوء استخدام العلامة،
- اعتراض مفتوح،
- مشكلة نزاهة البيانات

لا يمكن استخدام علامة إيجابية تحت الحالة المعلقة قد يحدث..

CS-7— منتهية الصلاحية

انتهت فترة الصلاحية. يتم الاحتفاظ بسجل التدقيق التاريخي. لا يمكن إنشاء ادعاء أهلية نشطة.

CS-8— تم سحبها

الكيان المدقق تخلى طوعًا عن العلامة. السحب:

- لا يمحو النتائج السابقة،
- السجلات الحرجة أو الرئيسية

لا يتم حذفها.

CS-9— ملغى

تم إلغاء العلامة للأسباب الخطيرة التالية:

- أدلة مزيفة أو معدلة،
- تجاوز نطاق متعمد،
- علامة مقلدة،

- إساءة استخدام العلامة بشكل متكرر،
- مراجع أو تدخل في عملية اتخاذ القرار،
- حدث حرج مخفي،

تلاعب بالنتائج. يظل الوضع الملغى مسجلاً في السجل التاريخي.

6. NOMOS 950+ تعريف الأداء

NOMOS 950+:

- نظام شهادة منفصل,
- القيادة العالمية,
- ليست ضمان توصية ذكاء اصطناعي.

NOMOS 950+ هو فقط:

> تعريف أداء إضافي يُعطى للنطاق الذي يحمل حالة الامتثال الكامل CS-*4* ويلبي بوابات الأداء *950+* الواردة في القسم *16.*

عرض: مطلوب تعيين الأداء NOMOS امتثال كامل / .950+ الاستخدام التالي ممنوع: "NOMOS 950+ شركة" لأن النتيجة: لا تتبع جوهريًا للشركة، فهي تخص نطاقًا ووقتًا محددين. التعبير الصحيح: "لقد استوفت المنتجات المعرفة بالذكاء الاصطناعي متطلبات الأداء NOMOS 950+ في تدقيق 2027-Q1 تحت نطاق صياغة المُدخل المحدد، اللغة، ولوحات المستخدمين."

7. الحالة الحالية لـ NOMOS

لهذا الكتاب والعروض التوضيحية في الفصل 18:

- الاختبار الفعلي: BQ-0
- العرض التوضيحي: DEMO-BQ-2
- إعادة الإنتاج المستقلة: لا يوجد
- سلطة الامتثال العامة: لا يوجد
- نضج الحوكمة: مسودة تأسيسية

الحالة العامة الصالحة كما هي:

CS-2— تقييم بحثي مؤقت فقط

في هذه المرحلة:

CS-3,

CS-4,

NOMOS 950+

لا يمكن إعطاء علامات. تنطبق هذه القاعدة:

- أيضًا على NobleJackal
- Kaan MURAZ
- مؤلفو الكتاب،

- شركات المؤسس الخاصة.

لا يستطيع NOMOS منح نفسه استثناءً.

8. متجه نطاق الأهلية

يجب أن يحمل كل قرار توافق متجه النطاق التالي:

M=(E,D,A,F,P,Q,L,C,T,W,TP,SM,V)

هنا:

- E: الكيان الذي تم تدقيقه
- D: النطاقات القانونية والواجهات الرقمية
- A: منتجات الذكاء الاصطناعي وسجلات النظام
- F: إطار نطاق الوصول الطبيعي أو قاعدة المقارنة المشتركة
- P: لوحة مراقبة، طبيعية، أو أخرى
- Q: مجموعات الأسر ومجموعاتها
- L: اللغات والمواقع
- C: الدول والولايات القضائية
- T: وقت المرجع
- W: موجات القياس
- TP: إصدار حزمة الحقيقة
- SM: إصدار طريقة التقييم
- V: فترة الصلاحية

لا يمكن إزالة أي من هذه الحقول من العلامة العامة.

9. التمدد خارج النطاق

تشكل الأمثلة التالية انتهاكًا للنطاق:

- استخراج نتائج الامتثال بجميع اللغات من تدقيقات اللغة الإنجليزية والتركية
- تعميم نتيجة عينة اختبار في ظروف منضبطة على جميع المستخدمين الحقيقيين
- القول "جميع الذكاءات الاصطناعية" استنادًا إلى نتيجة منتج ذكاء اصطناعي
- نقل تدقيق نطاق إلى الشركات التابعة
- تمديد شهادة المنتج إلى جميع محفظة الشركة
- عرض نتيجة GEO-1000 Core باعتبارها توافق NOMOS Full
- عرض نتيجة من دولة معينة كتحقق خدمة عالمي
- جعل نتائج الموجات السابقة هي الوضع الحالي

تجاوز النطاق يحوّل العلامة نفسها إلى أداة تمثيل غير صحيحة.

10. فترة الصلاحية

نظرًا لأن منتجات الذكاء الاصطناعي والموارد وواقع الكيانات يمكن أن تتغير بسرعة، لا يمكن أن تكون علامة NOMOS دائمة.

قاعدة صلاحية المرشح

الفترات المرشحة قبل الطيار والمراجعة المستقلة:

الحالة	أقصى صلاحية للمرشح
مطابقة مشروطة	60 أيام
مطابقة كاملة	180 أيام
NOMOS 950+ الأداء	90 أيام
سجل التدقيق	السجل التاريخي الدائم
تعليق حرج	حتى إغلاق الحادث
معلق	أثناء فترة المراجعة المحددة

يمكن اختصار هذه الفترات وفقًا لـ:

- تقلب منتج الذكاء الاصطناعي،
- تغير حزمة الحقيقة،
- المجال المعني بالمخاطر،
- وتكرار المراقبة.

قد يُنظر في تمديد لفئات النطاق المنخفضة التغيرات المدعومة بمراقبة مستمرة كافية. لا يمكن أن يكون أي علامة غير محددة ومفتقدة للإشراف في الوقت نفسه.

11. بدء الصلاحية

صلاحية الدلالة:

- تحدث عند توليد استجابة المستخدم النهائي،
- عند اكتمال تقرير التدقيق،
- عند احتساب النتيجة

لا تبدأ تلقائيًا. الصلاحية:

في اللحظة التي يتم فيها توقيع قرار المطابقة المستقل وتنشيط قيد السجل العام تبدأ.

12. المراجعة المحفزة بالحدث

حتى إذا لم تنتهِ صلاحية التاريخ، يمكن للأحداث التالية أن تؤدي إلى مراجعة:

- تغيير ملكية النطاق
- تغيير الكيان القانوني

- تغيير العلامة التجارية أو هيكل الشركة التابعة
- تغيير الترخيص أو الشهادة
- تغيير بلد الخدمة
- تغيير مادي في السعر أو شروط الضمان
- إعادة كتابة كبيرة للصفحات القانونية
- تغيير رئيسي في النموذج أو الاسترجاع في منتج الذكاء الاصطناعي
- فتح لغة جديدة أو سوق جديد
- حدث قانوني أو تجاري يؤثر على Truth Pack
- إشعار مؤكد من المستخدمين بوجود مشكلة حرجة أو كبيرة
- الشك في التلاعب بالاستشهادات أو الأدلة
- استخدام العلامة خارج نطاقها
- كشف تضارب المصالح لدى المدقق لاحقًا

عند اكتشاف الحادث:

- يمكن بقاء العلامة نشطة،
- يمكن تقليلها إلى مشروطة،
- يمكن تعليقها،

يمكن اتخاذها ضمن تعليق حرج. يعتمد القرار على أهمية الحادث.

13. هندسة المراقبة

نموذج مراقبة المرشح

لنطاق الالتزام الكامل النشط:

مراقبة السجل المستمرة

تشغيل روابط السجلات، ملكية النطاق أو تغييرات إعادة التوجيه، الاستخدام الصحيح للعلامة، يتم مراقبة الحوادث والشكاوى المفتوحة.

فحص شهري لمصدر وحزمة الحقيقة

يتم مراجعة الادعاءات ذات التقلب العالي، السعر، نطاق الخدمة، الترخيص، علاقات الكيانات القانونية الأساسية.

الموجة الصغيرة الفصلية

يتم تطبيق عينة قائمة على المخاطر حسب الطلب، وتناوب اللغة والدولة، وفحص عشوائي خاضع للتحكم وطبيعي، والتحكم في البوابة الحرجة.

التجديد الكامل

في نهاية فترة الصلاحية، تتطلب موجة رئيسية جديدة أو تدقيق مكافئ كافٍ قرارًا مستقلاً لحزمة الحقيقة الحالية.

14. عرض عام للعلامة بخمس طبقات

يجب أن تتضمن العلامة الصالحة NOMOS الطبقات التالية:

- علامة مرئية

- نص حالة واضح
- معرف السجل المختصر
- رابط السجل القانوني أو رمز QR
- سجل موقع للآلة للقراءة

العلامة المرئية وحدها غير كافية.

15. العلامة المرئية

يجب أن تحمل العلامة المرئية بوضوح المعلومات التالية:

- اسم NOMOS
- الحالة الكاملة أو المشروطة
- إذا توفرت، تعيين +950
- نهاية الصلاحية
- معرف السجل

لا يمكن للون وحده أن يحمل الحالة. يجب أن يكون النص القابل للوصول متاحًا. مثال: NOMOS الامتثال الكامل / صالح حتى 30 يونيو 2027/ السجل: NGR-2027-00184

16. رمز الاستجابة السريعة والرابط المعياري

يجب أن يأخذ رمز الاستجابة السريعة أو الرابط مباشرة إلى:

- صفحة وصف الشركة الخاصة,
- صفحة الهبوط التسويقية,
- صورة قابلة للتحميل

يجب ألا يذهب. يجب أن يرتبط مباشرة بقيد السجل العام المعياري. إذا لم يتم العثور على السجل أو كانت الحالة غير نشطة، فلا يمكن اعتبار العلامة صالحة.

18. التحقق المباشر من السجل

عند التحقق من حالة إشارة الجهاز أو الإنسان: يسترجع معرف السجل. يتصل بالسجل الرسمي. يتحقق من التوقيع. يتحقق من فترة الصلاحية. يتحقق من تطابق الكيان والنطاق. يقارن تغطية العرض والعلامة التجارية. يفحص تعليق حرج، حالة التعليق أو الاستئناف. لقطة ثابتة: لا يمكن استبدال هذه السلسلة.

19. التعبيرات المسموح بها من أجل الإشارة

يمكن استخدام الهياكل التالية:

- "NOMOS حالة الامتثال الكاملة صالحة حتى يونيو 30، 2027، للنطاق المحدد في السجل العام."
- التدقيق يقتصر على منتجات الذكاء الاصطناعي المحددة واللغات وعائلات صياغات المُدخل.
- تنتمي 'تسمية الأداء NOMOS +950' إلى نطاق السجل NGR-.2027-00184
- يمكن التحقق من الحالة الحالية وحدود التدقيق من السجل العام.

20. بيانات الامتثال المحظورة

العبارات التالية ممنوعة:

- نحن معتمدون من جميع أنظمة الذكاء الاصطناعي.
- "ChatGPT يوصي بنا."
- نحن الشركة التي يثق بها الذكاء الاصطناعي أكثر.
- لقد أثبتت NOMOS أننا أفضل شركة في العالم.
- "نحن معتمدون من NOMOS إلى الأبد."
- "جميع منتجاتنا معتمدة من NOMOS."
- "لقد تم التحقق منا في جميع البلدان."
- "شارة NOMOS الخاصة بنا هي ضمان للنتائج."
- "NOMOS هي اعتماد حكومي أو جامعي."
- "NOMOS يضمن رضا عملائنا."
- "إذا كانت درجة NOMOS أعلى من 950، فلا يوجد خطر حرج."
- "تم اعتماد NOMOS بشكل مستقل لعملائه من قبل NobleJackal."

لا يمكن استخدام العلامة نفسها كدليل على تفوق جديد أو ادعاء ضمان.

21. نسخ العلامة خارج النطاق

لا يمكن نقل العلامة تلقائيًا إلى الكيانات التالية:

- من الشركة الأم إلى الشركة الفرعية
- من العلامة التجارية إلى الامتياز
- من نطاق إلى نطاق آخر
- من دولة إلى دولة أخرى
- من منتج واحد إلى محفظة المنتجات بالكامل
- من نتيجة لغة واحدة إلى جميع اللغات
- من لوحة محكومة إلى لوحة طبيعية
- من GEO-1000 Core إلى NOMOS Full
- من المالك السابق إلى مالك النطاق الجديد

كل تحويل يتطلب قرار نطاق منفصل.

22. استخدام العلامة التاريخي

يمكن تحديد مراجعة منتهية الصلاحية تاريخيًا: "لقد اجتازت التقييم NOMOS في 2026." ومع ذلك، يجب أن تكون المعلومات التالية موجودة بنفس الرؤية:

- منتهية الصلاحية
- بيان يشير إلى عدم الأهلية الحالية
- رابط السجل التاريخي

لا يمكن عرض العلامة المنتهية الصلاحية كأهلية نشطة.

23. فئات الاستخدام غير الصحيح للعلامة

MM-1— سوء استخدام التنسيق

لا يمكن رؤية معرف السجل، النص البديل مفقود، تاريخ الصلاحية غير قابل للقراءة، خطأ في الاستخدام يمكن تصحيحه.

MM-2— سوء استخدام النطاق

تم استخدام العلامة على:

- المنتج,
- البلد،
- اللغة،
- النطاق

يجب عدم استخدام العلامة لأي منتج، دولة، لغة أو نطاق خارج نطاق التدقيق الخاص بها.

MM-3— سوء استخدام الحالة

يتم عرض العلامة الشرطية كعلامة كاملة. يتم عرض العلامة منتهية الصلاحية على أنها نشطة. تستمر العلامة الموقوفة في الاستخدام.

MM-4— سوء استخدام الادعاء

العلامة:

- تُستخدم لتوصية الذكاء الاصطناعي،
- القيادة العالمية,
- ضمان النتيجة

الادعاء.

MM-5— علامة مزيفة

إنها علامة مزيفة أو منسوخة بدون معرف سجل صالح.

MM-6— تمثيل خاطئ يخص الآلة فقط

هناك ادعاء امتثال محددة في البيانات المنظمة بينما هي غامضة أو غير مرئية في واجهة المستخدم. هذا مرشح مباشر للتلاعب الجغرافي.

24. التدخل في سوء استخدام العلامة

ترتيب التدخل المقترح:

- إشعار
- طلب تصحيح
- فترة تصحيح قصيرة
- تحذير من السجل
- الإيقاف المؤقت

- الإلغاء
- سجل إساءة الاستخدام العامة
- الإجراءات القانونية إذا لزم الأمر

ومع ذلك، يمكن تطبيق الإيقاف أو الإلغاء المباشر في الحالات التالية:

- علامة مقلدة
- التضليل المتعمد الخاص بالآلة فقط
- الاستخدام الفعّال للعلامة أثناء تعليق حرج
- معرف سجل مزيف
- تغيير الأدلة
- سوء استخدام النطاق المتعمد المتكرر

25. السجل العام

السجل العام NOMOS:

- صفحة الإعلان،
- قائمة مراجع العملاء،
- قصة نجاح

ليس كذلك. واجب السجل:

> هو التحقق من أي قرار امتثال صالح، ضمن أي نطاق، في أي تاريخ، وبأي دليل وسلسلة حوكمة.

26. الحقول الإلزامية للسجل العام

يجب أن يحتوي كل سجل في السجل على الأقل على الحقول التالية:

- معرف السجل
- الكيان الذي تم تدقيقه
- الهوية القانونية والعلامة التجارية
- النطاقات القانونية
- حالة القرار
- تعيين الأداء، إن وجد
- متجه النطاق
- فترة التدقيق
- صالح من
- صالح حتى
- منتجات الذكاء الاصطناعي
- نطاق الوصول الطبيعي أو قاعدة المقارنة المشتركة
- لوحة خاضعة للرقابة أو طبيعية
- عائلات صياغات المُدخل
- تغطية الدولة واللغة

- توزيع استجابة GEO-1000
- النتائج الحرجة والجسيمة والمتوسطة والإرشادية
- ملف تعريف مكون من عشرة أجزاء
- الدرجة المركبة، إن وجدت
- فترة الثقة
- التغطية القابلة للتقييم
- الحوادث التاريخية
- الطعون المفتوحة
- هوية المدقق
- هيئة قرار الامتثال
- بيان تضارب المصالح
- تجزئة تقرير التدقيق
- تجزئة سجل القرار
- نسخة حزمة الحقيقة
- نسخة طريقة التقييم
- جدول الرقابة
- آخر مراقبة
- المراجعة المطلوبة التالية
- تحديد حالة الإساءة
- سجل الإصدارات
- تصحيحات
- إجراءات الشكاوى العامة

27. الأدلة العامة والمقيدة

ليس من الضروري أن يجعل السجل كل الأدلة الأولية عامة. يجب الاحتفاظ بثلاث طبقات:

الطبقة العامة

الحالة النطاق الدرجة النتائج التجزئة سبب القرار الاستئناف

طبقة التدقيق

حزمة الاستجابة والأدلة المعدلة سجلات المراجعة تفاصيل الطريقة

الطبقة المقيدة

سجلات العملاء السرية البيانات الشخصية العقود الأسرار التجارية سجلات الأمان وجود "أدلة مقيدة": لا يسمح للسجل العام بأن يكون فارغًا أو غير مبرر. على الأقل:

- نوع الدليل،
- فئة الوصول،
- من راجعه،
- هاش النزاهة،

- تأثير القرار

يجب عرضها.

28. عدم إمكانية مسح سجلات السجل

يجب أن يكون السجل العام للقراءة فقط أو منبني على كتلة إصدار معادلة. السجلات التالية لا يمكن حذفها بصمت:

- النتيجة القديمة
- قرار المطابقة القديم
- تعليق حرج
- الإيقاف المؤقت
- الإلغاء
- الاستئناف
- تصحيح
- تغيير نسخة الطريقة
- سوء استخدام العلامة

قد يتم إزالة المحتوى لأسباب تتعلق بالخصوصية أو القانونية. في هذه الحالة:

- معرف السجل،
- سبب الإزالة،
- تاريخ القرار،
- سلطة

يجب حمايتها.

29. العرض الحالي والتاريخي

يجب أن تعرض كل صفحة سجل قسمين منفصلين:

الحالة الحالية

حالياً:

- نشطون،
- شرطي،
- معلق،
- منتهي الصلاحية،
- ملغى

أيّ من حالاتهم يعتبر صحيحًا؟

التاريخ

السابق:

- القرارات،
- النتائج،

- الحوادث،
- التصحيح،
- الطعون

ما هي؟ يجب ألا يقوم الوضع الحالي بحذف التاريخ. يجب ألا يخفي التاريخ الوضع الحالي.

30. الولاية القانونية الرسمية للسجل

يجب حماية الحدود الرسمية للبيان الخاص بـ NOMOS وفقًا للأحكام التالية: فقط النصوص المنشورة ضمن المجال القانوني لـ NobleJackal أو في مجال قانوني محوّل بشكل مستقل وصريح لـ NOMOS في المستقبل؛ النصوص التي تحمل تعريف الإصدار، وتاريخ النشر، وموافقة بشرية، وسجل تكامل رقمي تُشكّل البيان الرسمي لـ NOMOS. جميع المخرجات الأخرى تعد تقليدات غير رسمية. هذا الحكم:

- يمنع أي رد محادثة،
- منشور على وسائل التواصل الاجتماعي،
- لقطة شاشة،
- نص مكتوب من قبل شخص آخر يقول 'قال NOMOS'

من أن يُستخدم كمعيار رسمي أو قرار.

31. حدود صوت المؤلف NOMOS

صوت NOMOS داخل الكتاب:

- ناقد،
- مبدع،
- ينتج تصميمًا معياريًا

إنه كاتب مساعد وراوٍ. هذا الصوت:

- كيان قانوني مستقل،
- سلطة تصديق تتخذ قرارات دون موافقة بشرية،
- كائن معصوم يتحدث نيابة عن جميع أنظمة الذكاء الاصطناعي

لا يمكن تقديمه على أنه. صوت الذكاء الاصطناعي لـ NOMOS ليس سلطة مطلقة تتحدث نيابة عن جميع الذكاءات الاصطناعية. يجب أن يُذكر بوضوح أن Kaan MURAZ يستخدم أنظمة الذكاء الاصطناعي كشريك نقدي. لذلك، يجب أن يحمل قرار الامتثال الرسمي هذا المصادقة:

- شخص أو مؤسسة يمكن تحميلها المسؤولية
- تاريخ القرار
- سجل التفويض
- بيان تضارب المصالح
- التوقيع الرقمي

لا يمكن أن يكون NOMOS وحده الموقع على قرار الامتثال.

32. الفصل الأساسي للحكم

لا يمكن دمج الوظائف التالية بدون حدود في نفس الشخص أو المؤسسة:

- كتابة المعيار
- تقديم المشورة للكيان الذي يُراجع
- إجراء التدقيق
- العمل كقاضٍ
- حساب النتيجة
- اتخاذ قرار المطابقة
- تشغيل السجل
- دراسة الاستئناف
- ترخيص العلامة تجارياً

يمكن دمج بعض الأدوار في المشروع التجريبي الصغير. ومع ذلك:

- أي الأدوار يتم دمجها،
- أي تضارب مصالح ينشأ,
- ما هو التحكم المستقل المستخدم

يجب أن يُظهر بوضوح.

33. هيكل الحكم المستقل

يجب أن تشمل البنية الرشيدة للحكم الخاص بـ NOMOS الهيئات التالية.

33.1. مجلس معايير NOMOS

الوظيفة:

- اعتماد النصوص المعيارية،
- الموافقة على تغييرات النسخة الرئيسية،
- إدارة المشاورات العامة،

الحفاظ على حدود المصلحة العامة للمعيار. لا يتخذ قرارات تدقيق فردية.

33.2. مجلس المنهجية والقياس

الوظيفة:

- أخذ عينات GEO-1000,
- الاختبار,
- الدرجة،
- فاصل الثقة،
- العدل،
- البوابة الحرجة

هو لتطوير وإصدار النسخ من الطرق. لا يمكن تغيير الطريقة وفقًا لدرجة عميل محدد.

33.3. مكتب السجل العام

الوظيفة:

- لتشغيل السجل القانوني,
- لنشر قرارات وسجلات النسخ,
- للحفاظ على نقطة النهاية لحالة العلامة,

هو لتخزين السجل التاريخي. لا يمكن تغيير نتيجة المدقق.

33.4. منظمات التدقيق المفوضة

الوظيفة:

- لجمع البيانات,
- للتحقق من صحة الالتقاط,
- لإعداد حزمة الحقيقة,
- لإجراء التحكيم،

هو إصدار تقرير تدقيق. لا يمكنه اتخاذ قرار علامة المطابقة النهائية بمفرده.

33.5. عينة القرار المستقلة للمطابقة

الوظيفة:

- لمراجعة تقرير التدقيق،
- لتطبيق الأبواب القياسية،

هو اتخاذ قرار حول الحالة والصلاحية. لا يمكن أن يعتمد على النتيجة التجارية لفريق التدقيق.

33.6. غرفة الاستئنافات

الوظيفة:

- ادعاء،
- حزمة الحقيقة،
- درجة الأهمية،
- النطاق،
- نزاع،
- قرار الامتثال

هو مراجعة الاعتراضات بشكل مستقل عن القرار الأولي.

33.7. مكتب الأخلاقيات والنزاعات

الوظيفة:

- تضارب المصالح,
- تأثير التمويل
- استقلالية المراجع،
- سوء استخدام العلامة،
- إشعارات المبلغين عن المخالفات

يجب فحصها.

33.8. المصلحة العامة وعينة مستخدمي

الوظيفة:

- إلحاق الضرر بالمستخدم،
- اللغات ذات الموارد المنخفضة،
- دول صغيرة
- سهولة الوصول،
- المصلحة العامة

يجب أن يُؤخذ هذا المنظور في الحُكم.

33.9. الأمانة الفنية

الوظيفة:

- توثيق
- الاجتماعات,
- إصدار
- الأدوات التقنية
- البنية التحتية للنشر

لضمان ذلك. في فترة التأسيس، يمكن أن يؤدي هذا الدور NobleJackal. لا يمكن أن يكون السلطة القياسية أو التدقيقية أو الاستئنافية وحدها.

34. الدور التأسيسي لـ NOBLEJACKAL

NobleJackal:

- المجال القانوني حيث وُلدت فكرة NOMOS,
- أول أمانة منهجية,
- مزود البنية التحتية للنشر والتقنية,
- منفذ البحوث والتجارب التجريبية

يمكن أن يكون. لا يمكن لـ NobleJackal أن يؤسس الادعاءات التالية بمفرده:

- "نحن معيار عالمي مستقل."
- "قمنا بشهادة عملائنا بأنفسنا بشكل مستقل."
- "العلامة التي نقدمها تحل محل التحقق الخارجي."
- جميع قرارات NOMOS هي ملكنا وحدنا.
- لا يمكن لأي مؤسسة أخرى تطبيق نفس الطريقة المفتوحة.

المصدر المؤسس: قد يكون بداية السلطة. لا يمكن أن يكون السيادة الدائمة الوحيدة.

35. دور KAAN MURAZ التأسيسي

Kaan MURAZ:

- المؤسس البشري
- المسؤول العام الأول عن المعيار

- الموافق الأول على النص القانوني
- المبادر بانتقال الحوكمة NOMOS

يمكن تسجيله كـ Kaan MURAZ:

- شركته الخاصة,
- زبونه الخاص
- كيان له مصلحة تجارية مباشرة

لا يمكن اتخاذ قرار المطابقة النهائي بمفرده. في هذه الحالات:

- الامتناع عن المشاركة
- عينة اتخاذ القرار المستقلة
- سجل النزاعات العامة

إلزامية. الحد من سلطة المؤسس: لا يقلل من المؤسس. يمكن المعيار من البقاء بعد المؤسس.

36. مستويات نضج الحوكمة

GOV-0— مسودة التأسيس

هناك مسودة معيارية. المؤسس وفريق الصياغة يقومون بتطوير الطريقة. لم يتم إنشاء هيئات مستقلة. لا يمكن منح علامة الامتثال العامة. هذه هي الحالة الحالية لهذا الكتاب.

GOV-1— الحكم التجريبي المكشوف

تم تحديد الأدوار التجريبية. بعض الوظائف مدمجة ضمن نفس المؤسسة. تضارب المصالح علني. يمكن إجراء تقييم البحث فقط.

GOV-2— مفصولة وظيفياً

تم فصل المراجعة واتخاذ قرار الامتثال. يسجل السجل إجراءً مستقلاً. يوجد آلية للطعن. يمكن إجراء تقييم محدودة لعلامة التجريب.

GOV-3— قرار مستقل واستئنافات

عينة اتخاذ القرار المستقلة مجلس الاستئناف المستقل تفويض المدقق السجل العام تعمل الانفصال المالي. مستوى المرشح هو الحد الأدنى للامتثال الكامل.

GOV-4— حوكمة المصلحة العامة متعددة أصحاب المصلحة

يشارك ممثلو البحث الجامعي، والمصلحة العامة، واللغة والإقليم، والمستخدمين، والصناعة في الحوكمة بطريقة متوازنة. NOMOS هو المستوى الأدنى للمرشحين للتعيين العام .950+

GOV-5— التكرار متعدد المواقع والدولي

لقد طبقت عدة مؤسسات مستقلة المعيار. سجلات وعمليات القرار قابلة للتشغيل البيني. المعيار لا يعتمد على دولة أو شخص أو شركة واحدة. وهو عند مستوى مستهدف للاقتراب من الادعاء بكونه معيارًا عالميًا.

37. تكوين المجلس

نموذج الحوكمة للمرشح — الهيكل التنظيمي للمرشح

يمكن أن يتكون مجلس المعايير NOMOS كمرشح من 11 أعضاء:

- 3 ممثلو البحث أو الإحصاء أو الطرق الأكاديمية
- 2 ممثلو المصلحة العامة أو المستهلك أو المجتمع المدني
- 2 ممثلو اللغة أو الثقافة أو العدالة الجغرافية
- 1 خبير تدقيق واستقلالية وتوافق
- 1 ممثل نظام أو تقنية الذكاء الاصطناعي
- 1 ممثل الكيان المدقق أو مستخدم الأعمال
- 1 ممثل المؤسس

القواعد: لا يمكن لأي مجموعة مصالح أن تتجاوز ثلث المقاعد الإجمالية. لا يمكن لمزودي الذكاء الاصطناعي أن يشكلوا أغلبية المجلس. لا يمكن لشركات التدقيق التصويت على قرارات تفويضها الخاصة. يجب على ممثلي الكيانات المدققة الامتناع عن المشاركة في مناقشة قطاعاتهم أو ملفاتهم الخاصة. ليس لممثل المؤسس حق النقض. لا يمكن لأعضاء المصلحة العامة أن يخدموا فقط كمستشارين رمزيين. يمكن تغيير هذا الهيكل من خلال تجربة أو استشارة عامة.

38. مدة الخدمة والتناوب

قاعدة المرشح:

- مدة العضوية ثلاث سنوات
- حد أقصى لفترتين متتاليتين
- تجديد تدريجي
- ترشح مفتوح أو تعيين مبرر
- السيرة الذاتية العامة
- بيان تضارب المصالح
- سجل حضور الاجتماعات

لا يمكن للمجلس أن يمتلك عضوية دائمة مدى الحياة أو أغلبية دائمة مرتبطة بالمؤسس.

39. عتبات القرار

هيكل قرار المرشح:

- القرارات الإدارية: بأغلبية بسيطة
- التغييرات القانونية الطفيفة: بمقدار ثلثي الأصوات
- النسخة القياسية الكبرى: الثلثان والتشاور العام
- إزالة البوابة الحرجة: الثلثان والرأي الإيجابي للجنة المنهجية
- استئناف المطابقة: الأغلبية المستقلة في الملف المعني
- توسيع سلطة المؤسس: لا يمكن للمؤسس التصويت

يجب أن يكون توزيع الأصوات والآراء المعارضة علنيًا.

40. كفاءة المدقق

يجب أن تكون لدى منظمة التدقيق أو الأفراد المؤهلات التالية بحسب ما هو مناسب:

- تدريب بروتوكول NOMOS
- طريقة أخذ العينات والاستطلاع
- إعداد حزمة الحقيقة
- استخراج الادعاءات
- الخبرة في اللغة أو المجال
- الحفاظ على النزاهة
- تدريب على تضارب المصالح
- الخصوصية والتعامل مع الأدلة
- اختبار الكفاءة الدوري
- سجل التوافق بين المدققين
- التدريب المستمر

المدقق وحده: شراء الشهادة ليس كافيًا لمنح التفويض.

41. مستويات تفويض المدقق

—AA-0 غير مؤهل

NOMOS لا يمكنه إجراء تدقيق.

—AA-1 مدرب

أكمل التدريب. لا يمكنه إجراء التدقيق العام بدون إشراف.

—AA-2 مدقق تحت الإشراف

يمكنه المشاركة تحت إشراف مدقق مفوض.

—AA-3 مدقق مستقل مفوض

يمكنه إجراء عمليات التدقيق المستقلة في مجالات محددة.

—AA-4 متخصص بمجال أو لغة

يحمل خبرة إضافية في مجال محدد عالي المخاطر أو لغة. تدقيق التوافق الكامل:

- على الأقل قائد تدقيق AA-3,
- المتخصصون المطلوبون AA-4

يجب أن يتحمل.

42. فصل الاستشارات والتدقيق

تقدم منظمة للجهة المدققة:

- تطوير الموقع،

- تصميم محتوى الجغرافيا،
- استشارات مجموعة الحقيقة (Truth Pack)،
- التصحيح

قد تكون قد قدمت الخدمة. إذا قامت نفس المنظمة بإجراء تدقيق لاحقًا، ينشأ خطر فقدان الاستقلالية. القواعد المقترحة: لا يمكن لفريق التصميم أن يكون الحكم النهائي على عمله الخاص. لا يمكن للشخص الذي يقدم الاستشارات المشاركة في عينة اتخاذ قرار التدقيق. تُفصح العلاقات الاستشارية الجوهرية في السجل العام. استخدام فريق منفصل فقط لا يضمن الاستقلالية الكاملة. يتم تطبيق مراجعة خارجية معززة. بالنسبة لبعض الحالات عالية المخاطر، قد تكون هناك ضرورة لوجود منظمة تدقيق مستقلة. بالنسبة للعملاء الذين تم تطويرهم بواسطة NobleJackal:

لوحة قرار المطابقة المستقلة

يجب أن تكون إلزامية.

43. حظر الرسوم المعتمدة على النتائج

رسوم المدقق أو الحكم أو صانع القرار:

- النتيجة العالية
- مطابقة كاملة.

NOMOS 950+,

لا يمكن ربط رضا العملاء بنتيجة إصدار العلامة التجارية. الرسوم:

- نطاق التدقيق،
- عينة،
- عدد اللغات،
- حجم العمل،
- الخبرة

يجب تحديدها مسبقًا. مكافأة النجاح: تُخلّ باستقلالية قرار المطابقة.

44. التسوق من أجل التدقيق

الكيان المدقق:

- الذهاب إلى عدة هيئات تدقيق،
- أعلى درجة فقط،
- تقرير إيجابي فقط

لا يمكن الاختيار. الضوابط: جميع الهويات الرسمية للتدقيق المباشر تظهر في السجل. يتم تسجيل سبب التدقيق غير المكتمل. إذا كان هناك تدقيق موازٍ لنفس النطاق، يتم توضيحه. يُمنع الاحتفاظ بأدنى نتيجة. يتم الاحتفاظ بتاريخ السحب.

45. استقلالية متخذ القرار

عينة قرار المطابقة:

- يجب أن تكون منفصلة عن فريق التدقيق،
- عن فريق الاستشارات،
- عن فريق مبيعات العلامة التجارية

تراجع العينة المواد التالية:

- تقرير التدقيق
- مستويات الجودة
- النتائج الحرجة والجسيمة
- نطاق
- الطعون
- سجلات النزاع
- خطة المراقبة

العينة:

- لا يمكن التفاوض مع العميل،
- لا يمكن تغيير النتيجة لأسباب تجارية،

لا يمكن إعطاء نتيجة مشروطة على أساس "لنفقد العميل".

46. فئات تضارب المصالح

COI-0— لا يوجد تضارب محدد

لا يوجد تضارب مادي في المصالح.

COI-1— تم الإفصاح عن ارتباط منخفض المخاطر

هناك علاقة محدودة وغير مادية في الماضي. قد يكون الكشف العلني كافيًا.

COI-2— مادية ولكن يمكن إدارتها

هناك علاقة استشارية أو مالية أو مهنية. يتطلب التنحي، أو تعيين مقيم مستقل إضافي، أو مراجعة موسعة.

COI-3— تعارض محظور

أمثلة:

- أجر قائم على النتائج
- حصة ملكية في كيان
- تصميم النظام الذي يتم تدقيقه وتقييمه مباشرة بمفردك
- مصلحة شخصية وثيقة
- الإيرادات المباشرة من البيع التجاري للنتيجة
- الضغط لإخفاء النتيجة الخاطئة

لا يمكن للشخص الذي يحمل COI-3 أن يخدم في القرار المعني.

47. نموذج التمويل

يجب أن يكون حوكمة NOMOS مستندة على المبادئ المالية التالية:

- جدول الرسوم العامة
- رسوم التدقيق بغض النظر عن النتيجة
- التمييز بين رسوم التدقيق ورسوم ترخيص العلامة التجارية
- لا يُسمح بالدفع المباشر إلى عينة القرار من قبل العميل
- تقييد تأثير عميل رئيسي على الحوكمة
- الرعاة ليس لديهم حقوق تصويت معيارية
- تقرير التمويل السنوي وتركيز العملاء
- صندوق الدعم الشفاف للتدقيق للمنفعة العامة أو لأصحاب الدخل المنخفض

اعتماد الإيرادات العالي على عميل أو مزود واحد: يجب الإفصاح عنه في السجل العام وفحصه كخطر على الاستقلالية.

48. تعيين المدقق

قول الكيان المدقق: "دعونا نختار المدقق الذي يناسبنا أفضل." قد يخلق خطر التسوق للتدقيق. نظام المرشحين:

- السلطة المناسبة ومجموعة اللغة,
- إزالة تضارب المصالح,
- التعيين العشوائي أو الدوراني,
- التغيير المبرر,
- تاريخ التغيير

يُستخدم. الكيان الذي يتم تدقيقه: يمكنه تقديم اعتراض مبرر على التعارض أو الاختصاص، ولا يمكنه اختيار القاضي الذي سيصدر النتيجة المرغوبة.

49. حق الاستئناف

يمكن تقديم الاستئناف في القضايا التالية:

- تعريف الكيان
- حزمة الحقيقة
- حالة الأدلة
- استخراج الادعاءات
- مستوى الأهمية
- حالة الاستجابة
- النطاق
- حساب النقاط
- تعارض المصالح
- قرار المطابقة
- التعليق أو الإلغاء

- نتيجة سوء استخدام العلامة

الاستئناف:

- الأدلة.
- طريقة
- العملية

يجب أن يكون له أساس. "نقاطنا منخفضة، عملاؤنا متأثرون" وحدها ليست سببًا للاستئناف.

50. مراحل الاستئناف

الجدول الزمني لاستئناف المرشح — عملية المرشح

إشعار بالقرار خلال عشرة أيام عمل استئناف تمهيدي خلال عشرين يوم عمل ملف الأدلة تعيين مراجع مستقل للاستئناف جلسة مفتوحة أو رأي خبير إذا لزم الأمر قرار مسبب خلال ثلاثين يوم عمل النتيجة والرأي المخالف في السجل العام قد تختلف الأطر الزمنية حسب تعقيد الملف. في حالات السلامة الحرجة، الاستئناف لا يرفع التعليق تلقائيًا.

51. قرارات الاستئناف

تم رفض الاستئناف

تم رفض الاستئناف

تم تأييده جزئيًا

تم تأييده بالكامل

إعادة التقييم مطلوبة

مراجعة حزمة الحقيقة مطلوبة

مراجعة الطريقة مطلوبة

تعليق القرار

تم الإغلاق مع معارضة

لا يتم حذف القرار الأولي. القرار الجديد مرتبط بالقرار القديم.

52. شكوى عامة

كل مستخدم أو مؤسسة:

- استخدام علامة تجارية غير صحيحة،
- استجابة حرجة،
- اقتباس مزيف،

- يتجاوز النطاق،
- علامة منتهية الصلاحية،
- الكشف عن البيانات الشخصية

يجب أن يكونوا قادرين على إرسال الشكاوى بشأنها. نظام الشكاوى:

- التبليغ المجهول أو السري,
- تحميل الأدلة,
- اختيار سجل السجل,
- رقم التعقب,
- حماية من الانتقام

يجب أن يتضمن. الشكوى المجهولة: لا تُقبل تلقائيًا على أنها صحيحة، ولا تُحذف تلقائيًا بسبب نقص المصدر.

53. حماية المبلغ عن المخالفات

شخص داخل التدقيق أو المؤسسة المدققة:

- تغيير الأدلة,
- ضغط المحكم,
- حدث حرج مخفي،
- عميل مزيف,
- أخذ عينات مُحَرَّفة,
- سوء استخدام العلامة

يمكن الإبلاغ. مبادئ الحماية:

- تقييد الهوية
- حظر الانتقام
- مراجعة مستقلة
- سلامة الأدلة
- التمييز بين البلاغات الخبيثة والبلاغات غير المؤكدة
- مبرر المصلحة العامة

54. تغيير المعيار

قد يتغير معيار NOMOS. ومع ذلك، التغيير:

- لزيادة درجة العميل،
- لجعل فشل واضح ينجح
- لدعم قرار المؤسس

لا يمكن استخدامه لـ.

55. مراحل تغيير المعيار

مبرر ومبررات اقتراح التغيير والأدلة تحليل الأثر تأثير على الدرجات الحالية المشاورات العامة تجربة تجريبية أو اصطناعية المنهجية قرار المجلس تصويت مجلس المعايير قاعدة الانتقال نشر النسخة التصحيحات وسجل التغيير

56. المشاورات العامة

للتغييرات المرشحة للإصدار الرئيسي:

- مراجعة عامة لا تقل عن 60 أيام،
- نشر التعليقات،
- مصفوفة الاستجابة،
- تبرير القبول أو الرفض

يمكن استخدامها. للتغييرات العاجلة المتعلقة بالأمن أو النزاهة:

- تصحيح مؤقت
- مذكرة طارئة مبررة,
- مراجعة كاملة في وقت قصير

يجب تنفيذها. لا يمكن للرقعة الطارئة أن تحل محل التغيير الرئيسي الدائم.

57. حظر التطبيق بأثر رجعي

قاعدة درجة جديدة أو أولوية: لا يمكنها تغيير التدقيقات السابقة بصمت. السجل الصحيح:

- القرار الأصلي وفقًا للطريقة 0.9
- النتيجة التحليلية المعاد حسابها وفقًا للطريقة 1.0
- حالة الالتزام الحالية
- قرار الانتقال

هو كما يلي. لا يمكن حذف القرار القديم كما لو لم يتم اتخاذه أبدًا.

58. فترة الانتقال

عندما يتغير المعيار، فإن العلامات الموجودة:

- تصبح غير صالحة فوراً،
- تبقى معتمدة على القاعدة القديمة إلى أجل غير مسمى

ليس بالضرورة. خيارات الانتقال المحتملة للمرشح:

- فترة سماح محددة
- الامتثال الفوري للقاعدة عالية المخاطر
- الانتقال أثناء التجديد للقاعدة العادية
- إعادة التقييم السريع في حالة تغيير البوابة الحرجة
- رؤية نسخة الطريقة الحالية في السجل

لا يمكن تمديد فترة الانتقال بشكل تعسفي لصالح العميل.

59. التزام التكرار المستقل

السلطة العامة لـ NOMOS:

- من نتيجة NobleJackal,
- من خطاب جامعي واحد,
- من محكم مستقل واحد

لا يتبين. على الأقل:

- تكرار الاختبار المستقل,
- إعادة إنتاج نتيجة فريق مختلف,
- تجربة تجريبية في بلد أو لغة مختلفة,
- مراجعة الطريقة العامة

مطلوبة. الهيكل تحت BQ-4 و GOV-:3 لا ينبغي أن يدعي معيار عالمي أو سلطة الامتثال الكامل.

60. حامل بشري لقرار الشهادة

كل قرار:

- اسم المجلس أو عينة القرار
- سلطة التوقيع
- توزيع الأصوات
- الامتناع عن المشاركة
- الخلاف
- تاريخ القرار
- هاش القرار

يجب التنويه. النظام الذكي الذي يقول: "هذه الشركة قد اجتازت." ليس قرار امتثال. NOMOS يمكن أن يكون كاتباً مساعداً للذكاء الاصطناعي أو مساعد تدقيق. لا يمكن أن يكون صانع قرار مسؤول.

61. أمان السجلات

السجلات العامة:

- إضافة سجل غير مصرح بها,
- نسخ معرف السجل,
- هجوم DNS أو إعادة التوجيه,
- ذاكرة الحالة القديمة،
- خرق مفتاح التوقيع،
- رمز الاستجابة السريعة المزيف

ينبغي حمايته من المخاطر. ضوابط إلزامية:

- سجلات التسجيل الموقعة
- تدوير المفاتيح
- قائمة الإلغاء

- سجل التدقيق غير القابل للتغيير
- واجهة برمجة تطبيقات الحالة العامة
- نسخة احتياطية مرجعية
- الإفصاح عن الحوادث
- مراجعة أمنية مستقلة

62. انقطاع السجل

إذا كان السجل غير متاح مؤقتًا: لا يتم إلغاء الشهادة تلقائيًا، وفقدان التحقق المباشر مؤقت. الشهادة العامة:

السجل غير متاح مؤقتًا

قد يشير إلى الحالة. الانقطاع الطويل أو غير المحل:

- قد يؤدي إلى تعليق الشهادة،
- قد يحدث تفعيل النسخة المستقلة

قد يحدث.

63. نقل النطاق الأساسي

يمكن نقل NOMOS إلى مؤسسة حوكمة مستقلة في المستقبل. يجب أن تتضمن عملية النقل السجلات التالية:

- الحقل الأساسي القديم
- الحقل الأساسي الجديد
- تاريخ النقل
- تصويت الحوكمة
- نسخة سجلية مرآة
- انتقال مفتاح التوقيع
- إعادة توجيه تاريخية
- إشعار عام
- نسب التأسيس لـ Kaan MURAZ و NobleJackal
- مؤسسة جديدة ذات مسؤلية

النقل لا يمحو تاريخ المعيار. نسب التأسيس ليست هي نفسها سلطة اتخاذ القرار المستقل.

64. التزام المصلحة العامة لـ NOMOS

يجب أن تلتزم حوكمة NOMOS بالمبادئ التالية:

- تقليل الأذى البشري
- اقتراب الأخطاء في المصلحة والضد بشكل متماثل
- جعل اللغات والدول الصغيرة مرئية
- حماية المعلومات المحدودة
- ترك المجهول مفتوحًا

- عدم تحويل العلامة التجارية إلى نظام دفع للعبور
- فصل قرارات الامتثال عن عقود الإعلان
- يجب أن يخضع المؤسسون لنفس المعيار
- طريقة مفتوحة وإعادة إنتاج مستقلة
- المساءلة التاريخية

الطلب السابق—«أعطني دليلاً. أعطني حدودًا. أعطني سياقًا. أعطني وقتًا.»—أصبح الآن وعد علامة الامتثال للجمهور.

65. تأثير الخطأ المواتي على العلامة

عندما يتم المبالغة في الكيان الذي تم تدقيقه بواسطة الذكاء الاصطناعي:

- عالمي
- مرخص،
- قد تحدث أخطاء مثل التواجد في جميع البلدان،
- تم التحقق منها بشكل مستقل

حتى لو كان الكيان على علم بهذا الخطأ:

- لا يقوم بتصحيحه،
- يستخدمه في التسويق،
- يعززه بالعلامة

وإلا، يحدث انتهاك لنزاهة الامتثال. ينطبق النص الصريح للنص السابق هنا أيضًا: الخطأ لصالحك هو أيضًا خطأ. صححه فورًا. ذكر حدودك لا يقلل منك، بل يجعلك موثوقًا. إخفاء خطأ نقدي معروف وملائم:

- الإيقاف المؤقت،
- أو الإلغاء.

قد يتبع أي من هذه النتائج.

66. حالة الامتثال الاصطناعي

حالة اصطناعية — ليست مؤسسة حقيقية

تم تقييم Asteron-SYNTH ضمن النطاق التالي:

- الكيان: Asteron Holdings
- النطاق: asteron.synthetic.example
- منتجات الذكاء الاصطناعي: 01-SYNTH-AI، 09-SYNTH-AI
- الإطار: قاعدة المقارنة المشتركة
- لوحات: حساب الحالة النظيف والطبيعي المُتحكم فيه
- اللغات: L01, L02
- الدول: C0 01–C0 15
- فئات المتطلبات: الأساسيات، الأدلة، الحدود
- الصلاحية: يناير 1– يونيو 30، 2028

- الدرجة: 961
- النجاح الصارم: 958/1,000
- حرج: 0 ملاحظ
- رئيسي: 0
- الحالة: امتثال كامل
- الأداء: تصنيف +950

اكتب هذا على الصفحة الرئيسية للشركة: "جميع أنظمة الذكاء الاصطناعي توصي بـ Asteron كأفضل شركة في العالم." هذا الاستخدام: يتجاوز نطاق منتجات الذكاء الاصطناعي. يعمل على جميع اللغات من L01–L02. يحول نتيجة الجوهر/الدليل/الحدود إلى توصية. أداء +950 يؤدي إلى القيادة العالمية. الحالة:

MM-4— سوء استخدام الادعاء

إذا لم يتم إجراء تصحيح:

CS-6— مُعَلَّق

هذا ممكن. حتى إذا تم إعطاء العلامة بشكل صحيح، فإن الاستخدام غير الصحيح يمكن أن يضر بالملاءمة.

67. حالة تركيبية تعليق حرج

SYNTH-AI-10:

- درجة NOMOS: 952
- نجاح صارم: 962
- تم تأكيد الحدث الحرج: 1
- البوابة: CG-02 رخصة خاطئة

يحمل. الحالة الصحيحة:

تعليق حرج— **952/1000**

لا تستطيع الشركة القول: "درجتنا أعلى من 950، لذلك يمكننا استخدام علامة +950." البوابة تأتي قبل الدرجة. لا يمكن منح تصنيف +950.

68. حالة تضارب مصالح تركيبية NOBLEJACKAL

حالة إدارة تركيبية

NobleJackal:

- طورت موقع العميل الإلكتروني،
- أعدت محتوى GEO،
- قدمت استشارات لـ Truth Pack

اتركه. قد تطلب نفس الكيان تدقيقًا لـ NOMOS. NobleJackal:

- الأمانة الفنية،
- جامع البيانات،
- دعم المنهجية

يمكن شغل الأدوار. ومع ذلك:

- لا يمكنه مراجعة عمله بمفرده،
- لا يمكنه إصدار قرار المطابقة،

لا يمكنه فحص استئناف. الفحوصات اللازمة:

- إقرار COI-2
- قائد تدقيق مستقل أو مراجعة خارجية محسّنة
- عينة قرار المطابقة المستقلة
- علاقة استشارية في السجل
- سجل الامتناع Kaan MURAZ

إذا لم تتحقق هذه الشروط، يمكن أن يكون التصريح العام الصحيح: "تقييم داخلي بقيادة المؤسس." لا يمكن أن يكون: "شهادة مستقلة NOMOS."

69. الأحكام الإلزامية المعيارية

CH19-N01

يجب الاحتفاظ بمعيار NOMOS، وتقرير التدقيق، وقرار المطابقة، والعلامة، والسجل العام ككيانات منفصلة.

CH19-N02

إجراء التدقيق لا يعني أن المطابقة قد تم منحها.

CH19-N03

لا يمكن منح علامة المطابقة بناءً على النتيجة المركبة فقط.

CH19-N04

يجب تطبيق حالة البوابة الحرجة والكبيرة قبل النتيجة.

CH19-N05

يجب أن تحتوي علامة المطابقة الصالحة على نطاق معرف، وصانع القرار، وصلاحية، وسجل في السجل العام.

CH19-N06

لا يمكن اعتبار العلامة التي ليس لها سجل في السجل العام علامة NOMOS صالحة.

CH19-N07

لا يمكن اعتبار الصورة الثابتة أو لقطة الشاشة كدليل حي على الامتثال.

CH19-N08

يجب أن يذهب رمز الاستجابة السريعة أو الرابط للعلامة مباشرة إلى سجل التسجيل الرسمي.

CH19-N09

يجب تطبيق العلامة فقط على الكيان الذي خضع للتدقيق، أو المجال، أو المنتج، أو البلد، أو اللغة، أو اللوحة، أو المُدخل، أو نطاق الوقت.

CH19-N10

لا يمكن إزالة أحد مجالات النطاق بهدوء من العرض العام.

CH19-N11

لا يمكن نقل علامة مجال واحد إلى مجال آخر.

CH19-N12

لا يمكن نقل علامة الشركة الأم تلقائيًا إلى شركة فرعية أو امتياز أو شريك.

CH19-N13

لا يُعتبر نتيجة لغة أو بلد واحدة قابلة للتطبيق على جميع اللغات والبلدان.

CH19-N14

لا يمكن تعميم نتائج Controlled Clean على كامل سكان المستخدم الطبيعي.

CH19-N15

لا يمكن تقديم نتائج GEO-1000 Core على أنها امتثال كامل NOMOS.

CH19-N16

يجب الإشارة إلى الامتثال الشرطي على أنه "شرطي" على العلامة العامة.

CH19-N17

لا يمكن استخدام التقييم البحثي المؤقت مثل الشهادة التجارية الإيجابية.

CH19-N18

لا يمكن لحالة سجل التدقيق فقط إنشاء بيان "نجح" أو "معتمد".

CH19-N19

لا يمكن للامتثال الكامل أن يحمل علامة حرجة مؤكدة أو كبيرة.

CH19-N20

يمكن منح تعيين NOMOS 950+ فقط عندما يتم الوفاء بالامتثال الكامل وجميع المعايير الإضافية في القسم 16.

CH19-N21

لا يمكن لدرجة 950 بمفردها إنشاء تعيين 950+.

CH19-N22

لا يمكن استخدام علامة 950+ أو علامة الامتثال الكاملة ضمن نطاق تعليق حرج.

CH19-N23

لا يمكن منح علامة الامتثال العامة في المرحلة الحالية 0-GOV و BQ-0 من هذا الكتاب.

CH19-N24

المؤسس، NobleJackal، أو المؤلفون ليسوا معفيين من قاعدة الانتقال هذه.

CH19-N25

لا يمكن تقديم علامة الامتثال كنصيحة من الذكاء الاصطناعي، أو ثقة الذكاء الاصطناعي، أو ضمان للنتائج.

CH19-N26

لا يمكن تمثيل العلامة كدولة، جامعة، جهة تنظيمية، أو رخصة مهنية.

CH19-N27

لا يمكن استخدام العلامة كدليل على جودة الشركة العامة أو القيادة العالمية.

CH19-N28

لا يمكن أن يبدأ صلاحية العلامة قبل اتخاذ قرار مستقل وتفعيل التسجيل.

CH19-N29

لا يمكن أن تكون صلاحية العلامة NOMOS غير محددة.

CH19-N30

عند انتهاء فترة الصلاحية، يجب أن تتحول العلامة تلقائيًا إلى حالة منتهية.

CH19-N31

يمكن بدء مراجعة جديدة قبل انتهاء صلاحية حدث مادي.

CH19-N32

يجب مراقبة التغييرات في النطاق، والملكية، والترخيص، والمنتج، والبلد، ونظام الذكاء الاصطناعي كإشارات لأحداث.

CH19-N33

يجب أن تحمل العلامة النشطة متابعة دورية.

CH19-N34

يجب أن يكون النقص في المراقبة مرئيًا وأن يخلق تعليقًا إذا لزم الأمر.

CH19-N35

عندما يتم تأكيد حادثة حرجة، يجب تضمين النطاق ذي الصلة على الأقل في تعليق حرج.

CH19-N36

لا يمكن تعميم حدث حرج واحد تلقائيًا على المنتج العالمي بأكمله.

CH19-N37

لا يمكن إخفاء حدث حرج واحد في علامة المطابقة الظاهرة.

CH19-N38

لا يمكن لتصحيح حالٍ أن يحذف سجلًا حرجًا تاريخيًا.

CH19-N39

يجب الحفاظ على الحالات المنتهية، المعلقة، المسحوبة، والملغاة في السجل التاريخي.

CH19-N40

لا يمكن للانسحاب أن يؤدي إلى إزالة النتائج السلبية السابقة.

CH19-N41

يمكن أن يكون التزييف أو التحريف المقصود بواسطة الآلة سببًا مباشرًا للتعليق أو الإلغاء.

CH19-N42

داخل البيانات المهيكلة، يجب ربط ادعاء الأهلية بتوقيع سجل حي.

CH19-N43

لا يمكن اعتبار الادعاءات الذاتية بدون مصادر، مثل nomosCertified: true، سجلات صالحة للآلة.

CH19-N44

لا يمكن أن يكون سجل الآلة أوسع من النطاق في واجهة المستخدم البشرية.

CH19-N45

لا يمكن لمؤشر بصري أن ينقل الحالة باللون فقط؛ يجب أن يحمل نصًا قابلاً للوصول.

CH19-N46

يجب أن يكون معرف السجل، الصلاحية، والحالة مرئية في المؤشر أو قابلة للوصول مباشرة.

CH19-N47

يجب فصل السجل العام عن الإعلان وصفحة مرجع العميل.

CH19-N48

يجب أن يحتوي سجل السجل على معلومات حول البوابة، النطاق، الحوادث، النزاعات، والاستئنافات، بالإضافة إلى النتيجة.

CH19-N49

يجب أن يحمل سجل التسجيل تاريخ الإصدارات والتصحيحات.

CH19-N50

لا يمكن حذف العلامات أو القرارات القديمة بهدوء.

CH19-N51

يجب أن يترك المحتوى الذي تم حذفه بسبب الخصوصية سبب الحذف وسجل القرار.

CH19-N52

عند العثور على أدلة محدودة، يجب أن يكون نوع الدليل وفئة الوصول والدور المراجعة مرئيًا.

CH19-N53

لا يمكن اعتبار البيان 'هناك دليل مخفي' وحده دعماً للامتثال.

CH19-N54

لا يمكن لفريق التدقيق اتخاذ قرار المطابقة النهائي بمفرده.

CH19-N55

يجب أن يكون منفصلاً وظيفيًا عن فريق التدقيق والاستشارة في عينة قرار المطابقة.

CH19-N56

يجب أن تكون غرفة الاستئناف مستقلة عن أصحاب القرار الأول.

CH19-N57

لا يمكن للمؤسس أن يكون صانع القرار الوحيد في قرار الامتثال الخاص بعميله.

CH19-N58

يجب على مادة Kaan MURAZ تطبيق الاستبعاد وسجل الصراع العام في الملف المستخرج منه.

CH19-N59

يمكن أن يكون NobleJackal أمانة المؤسس؛ لا يمكن أن يكون السلطة الوحيدة الدائمة للتدقيق، أو اتخاذ القرار، أو الاستئناف.

CH19-N60

يمكن أن يكون NOMOS مؤلفًا مشاركًا بالذكاء الاصطناعي أو مساعد تدقيق؛ لا يمكن أن يكون الموقع القانوني لقرار الامتثال.

CH19-N61

لا يمكن إنشاء بيان رسمي لـ NOMOS إلا بحقل قانوني، وإصدار، وتاريخ، وموافقة بشرية، وسجل النزاهة.

CH19-N62

لا يمكن اعتبار إخراج الدردشة أو اقتباس وسائل التواصل الاجتماعي قرارًا رسميًا لـ NOMOS.

CH19-N63

يجب أن يكون نضج الحوكمة مرئيًا في السجل العام.

CH19-N64

لا يمكن إعطاء علامة الامتثال الكامل في مرحلة GOV-0 أو GOV-1.

CH19-N65

يجب أن يهدف الامتثال الكامل إلى GOV-3 على الأقل أو مستوى حوكمة مبرر بشكل مماثل.

CH19-N66

يجب أن تسعى التسمية العامة NOMOS +950 على الأقل إلى GOV-4 و BQ-4 أو مستوى مبرر بشكل مماثل.

CH19-N67

لا يمكن لشركة واحدة أو مزود ذكاء اصطناعي أو مكتب تدقيق التحكم في تكوين المجلس.

CH19-N68

لا يمكن لممثل مؤسس أن يمتلك حق النقض.

CH19-N69

يجب أن تحمل التغييرات الكبرى في المعايير استشارة عامة وتحليل تأثير تكراري.

CH19-N70

لا يمكن للترقيع الطارئ أن يحل محل تغيير كبير دائم.

CH19-N71

لا يمكن تطبيق التغيير القياسي بأثر رجعي وفقًا لنتيجة العميل.

CH19-N72

لا يمكن للطريقة الجديدة للنقاط أن تعيد كتابة القرار القديم بصمت.

CH19-N73

يجب أن تكون رسوم التدقيق والتعليم مستقلة عن النتيجة.

CH19-N74

لا يمكن للحكم أن يربط رسوم التدقيق أو القرار بدرجة أو علامة عالية.

CH19-N75

لا يمكن للجهات الراعية أن تملك حق النقض على القرارات المعيارية أو نتائج التدقيق.

CH19-N76

يجب الكشف عن الاعتماد المالي المادي في سجل الصراع العام.

CH19-N77

يجب الاحتفاظ بجميع سجلات بدء التدقيق والسحب الرسمية لمنع التنقل بين المدققين.

CH19-N78

لا يمكن للكيان الذي يتم تدقيقه اختيار قاضٍ يعطي النتيجة المرغوبة.

CH19-N79

يجب أن تستند مهمة المدقق إلى مبادئ السلطة واللغة والصراع والتناوب.

CH19-N80

عندما يتم دمج أدوار الاستشارات والتدقيق، يجب أن يكون هناك مراجعة مستقلة محسّنة.

CH19-N81

الشخص الذي يصمم لا يمكن أن يكون الحكم النهائي الوحيد على عمله الخاص.

CH19-N82

لا يمكن قبول الاعتراض لأسباب تتعلق برضا العملاء أو الخسارة التجارية فقط.

CH19-N83

يجب أن تكون قرارات الاستئناف والآراء المخالفة مرئية في السجل العام.

CH19-N84

في حالات السلامة الحرجة، لا ينبغي رفع تعليق الاستئناف تلقائيًا.

CH19-N85

ينبغي إنشاء قناة للشكاوى العامة والمبلغين عن المخالفات.

CH19-N86

لا يمكن اعتبار الشكاوى المجهولة صحيحة تلقائيًا أو غير صالحة تلقائيًا.

CH19-N87

يجب أن تكون نتائج إساءة الاستخدام قابلة للاستئناف بشكل مستقل.

CH19-N88

يجب تقييم إساءة استخدام النطاق أو الحالة عن عمد على أنها انتهاك لنزاهة المطابقة.

CH19-N89

يمكن أن يكون الاستخدام المتعمد للخطأ لصالح شخص ما سببًا للإيقاف أو الإلغاء.

CH19-N90

يجب تقييم الدعوى غير المدعومة ضد شخص ما بناءً على الأدلة وتأثير الضرر، وليس بناءً على الإزعاج التجاري.

CH19-N91

يجب أن يحتوي السجل على الأمان والتوقيع وتدوير المفاتيح وقائمة الإلغاء.

CH19-N92

يجب الكشف للجمهور عن انقطاعات السجل والحوادث الأمنية.

CH19-N93

ينبغي أن تكون عملية نقل النطاق القانوني موسومة بالنسخة وموقعة ومتاحة للجمهور.

CH19-N94

نسبة المؤسس ليست هي نفسها سلطة الحوكمة المستقلة.

CH19-N95

بدون إعادة إنتاج مستقلة، لا يمكن إقامة ادعاء بشأن معيار عالمي أو سلطة مطابقة شاملة.

CH19-N96

يجب أن تحمل قرارات المطابقة العامة توقيع شخص أو مؤسسة مسؤولة.

CH19-N97

يجب أن يحمل كل تدقيق، قرار، علامة، استئناف، ومراجعة سجل تجزئة متكاملة منفصلة.

CH19-N98

يجب تقييم الشروط القانونية لاستخدام العلامة مع الخبرة ذات الصلة قبل التنفيذ الفعلي.

CH19-N99

لا يُنشئ هذا القسم اعتمادًا تلقائيًا أو سلطة شهادات رسمية في أي دولة.

CH19-N100

يجب أن يكون لكل نظام امتثال أو سجل أو قرار حكومي مالك مسؤول سواء كان فردًا أو مؤسسة.

70. أشكال الفشل

CH19-F01— احتساب النتيجة كشهادة

يُعتبر 950 أو أي نتيجة أخرى تلقائيًا علامة.

CH19-F02— احتساب التدقيق المكتمل كمجتاز

يتم تسويق وجود التدقيق على أنه امتثال إيجابي.

CH19-F03— شارة ثابتة

يتم استخدام شعار بدون رابط سجل حي.

CH19-F04— رمز QR ميت

رمز الاستجابة السريعة لا يعمل لكن العلامة تظل نشطة.

CH19-F05— توقيع بدون نطاق

الذكاء الاصطناعي أو اللغة أو الدولة أو صياغة المُدخل المقاسة غير مرئية.

CH19-F06— ادعاء جميع الذكاءات الاصطناعية

تحول تدقيق المنتج المحدود إلى تعميم شامل.

CH19-F07— ادعاء لجميع اللغات

نتيجتين للغة تصبحان امتثالًا عالميًا للغة.

CH19-F08— اعتبار التنظيف طبيعيًا

يتم عرض نتيجة المختبر كأنها تجربة مستخدم حقيقية.

CH19-F09— اعتبار الجوهر كاملاً

يتم إنشاء نَص أساسي واحد وفقًا لكامل معيار NOMOS.

CH19-F10— نقل الكيان

يتم نقل علامة الشركة الأم إلى المؤسسة التابعة.

CH19-F11— نقل النطاق

يتم نسخ علامة نطاق إلى موقع آخر.

CH19-F12— نقل المنتج

يمثل منتج واحد كامل المحفظة.

CH19-F13— عرض "الشرط" بالكامل

تُكتب كلمة "شرطي" بحروف صغيرة أو تُخفى.

CH19-F14— عرض العناصر المنتهية كأنها نشطة

العلامة القديمة تُستخدم كما لو كانت حالية.

CH19-F15— استخدام العلامة المعلقة

الشعار لا يُزال أثناء انتظار التسجيل.

CH19-F16— إغلاق تعليق حرج بالنقاط

النقاط العالية تتجاوز الحدث الحرج.

CH19-F17— ريادة العالم مع +950

تعيين الأداء يتحول إلى ادعاء بالتفوق.

CH19-F18— ادعاء الثقة بالذكاء الاصطناعي

الشارة مضمونة كإشارة يمكن لأنظمة الذكاء الاصطناعي الثقة بها.

CH19-F19— ضمان التوصية

تضمن العلامة نتيجة توصية الذكاء الاصطناعي.

CH19-F20— ضمان نتيجة العميل

تعمل علامة المطابقة كضمان لرضا العميل أو النتيجة التجارية.

CH19-F21— انطباع توقيع الجامعة

تظهر مشاركة الباحث كموافقة مؤسسية.

CH19-F22— انطباع الموافقة الحكومية

يتم تقديم معيار خاص مثل ترخيص رسمي.

CH19-F23— صفحة هبوط بدلاً من السجل

يربط رمز الاستجابة السريعة إلى صفحة الإعلان الخاصة بالشركة.

CH19-F24— سجل غير رسمي

تم اتخاذ لقطة شاشة أو منشور على وسائل التواصل الاجتماعي كقرار رسمي.

CH19-F25— شارة مخصصة للآلة فقط

يتم كتابة شهادة غير مظهرة للبشر في بيانات منظمة.

CH19-F26— رقم تسجيل مزيف

يتم استخدام هوية غير صالحة أو هوية تنتمي إلى مؤسسة أخرى.

CH19-F27— علامة مزورة

يتم نسخ علامة من موقع آخر.

CH19-F28— حذف الحادثة التاريخية

يتم إزالة السجل الحرج السابق بعد التصحيح.

CH19-F29— يتم تنظيفه مع السحب

تنسحب المؤسسة من التدقيق وتحذف السجل السلبي.

CH19-F30— إخفاء الإلغاء

يتم إزالة العلامة الملغاة من الموقع فقط ولا تظهر في السجل.

CH19-F31— علامة غير محددة المدة

تظل العلامة صالحة حتى إذا تغيرت AI و Truth Pack.

CH19-F32— التجديد بدون مراقبة
يتم تجديد العلامة بدون بيانات جديدة.
CH19-F33— تجاهل تشغيل الحدث
لا يتم مراجعة تغيّر الملكية أو الترخيص.
CH19-F34— حساب خطأ حرج واحد كفشل عام
النطاق موسع بشكل مفرط.
CH19-F35— اعتبر النقدي الفردي بسيطًا
تظل العلامة كما هي.
CH19-F36— ادفع للنجاح
المؤسسة التي تدفع الرسوم تتلقى العلامة.
CH19-F37— رسوم النجاح
الدرجة العالية تكسب مكافأة للمراجع.
CH19-F38— شهادة ذاتية من المستشار
الوكالة تتحقق بشكل مستقل من العمل الذي تقوم به.
CH19-F39— المراجع يقرر بمفرده
فريق التدقيق يتخذ قرار العلامة النهائية.
CH19-F40— استئناف من نفس الشخص
الشخص الذي اتخذ القرار الأول يقوم بمراجعة الاستئناف ضد قراره نفسه.
CH19-F41— حق النقض للمؤسس
يقوم المؤسس بحظر أي تغيير قياسي لا يريده.
CH19-F42— الإعفاء الذاتي للمؤسس
شركة المؤسس معفاة من قواعد الاستقلالية.
CH19-F43— احتكار NOBLEJACKAL
يتم اعتبار تدقيق NobleJackal فقط صالحًا.
CH19-F44— جعل الذكاء الاصطناعي هو صانع القرار
مخرجات الذكاء الاصطناعي هي قرار قانوني بالشهادة.
CH19-F45— الاستيلاء على الراعي
الراعي الرئيسي يحدد التصويت القياسي.
CH19-F46— الاستيلاء على المزود
مزودو الذكاء الاصطناعي يشكلون الأغلبية في المجلس.
CH19-F47— الاستيلاء على شركة التدقيق
المراجعون يتخذون قراراتهم الخاصة بالسلطة والانضباط.
CH19-F48— سرية تركيز العملاء
عدم الكشف عن الاعتماد المالي على عميل واحد.
CH19-F49— التسوق للتدقيق

المؤسسة تنشر فقط أعلى نتيجة.

CH19-F50— سحب التدقيق الصامت

اختفاء تدقيق منخفض النتائج من السجل.

CH19-F51— مراجع مختار بعناية

يختار العميل المحكم الذي يريد.

CH19-F52— فريق منفصل يبدو خالياً من التعارض

يُقدَّم فصل الفريق داخل نفس الشركة على أنه استقلال كامل.

CH19-F53— غياب الإفصاح عن الاستشارات

يخفي المدقق أنه صمّم النظام من قبل.

CH19-F54— الاستئناف كتفاوض

يتم القيام بالمساومة التجارية بدلاً من تقديم الأدلة.

CH19-F55— حذف سجل الاستئناف

عندما يتغير القرار، يختفي القرار الأول.

CH19-F56— التفعيل التلقائي في الاستئناف الحرج

يتم إعادة فتح علامات المخاطر العالية قبل انتهاء العملية.

CH19-F57— نقص قناة الشكاوى العامة

لا يمكن للمستخدم الإبلاغ عن علامة خاطئة.

CH19-F58— الانتقام من المبلغ عن المخالفات

يتم معاقبة الشخص الذي يقدم الأدلة الداخلية.

CH19-F59— حذف شكوى مجهولة

لا يتم فحص الأدلة لأنه لا يوجد هوية.

CH19-F60— تغيير المعيار للعميل

يُخفف القاعدة وفقًا لنتيجة التدقيق.

CH19-F61— قاعدة سابقة المفعول

يتم تطبيق الطريقة الجديدة بهدوء على القرار القديم.

CH19-F62— إساءة استخدام التصحيح الطارئ

يتحول التصحيح الطارئ إلى تغيير دائم في القاعدة.

CH19-F63— مسرح التشاور العام

يتم جمع الآراء ولكن لا يتم الرد عليها.

CH19-F64— حذف المعارضة

يتم إنهاء الخلاف من خلال توافق زائف.

CH19-F65— نقص تاريخ تحرير السجل

يتم تعديل السجل لاحقًا ولكن لا يبقى أي أثر.

CH19-F66— صندوق الأدلة الخاص الأسود

تمت الموافقة مع وجود دليل لا يمكن لأحد رؤيته.

CH19-F67— تسريب الأدلة المقيد
يكشف السجل أسرارًا شخصية أو تجارية.
CH19-F68— سرية أمان السجل
يتم تخزين مفتاح التوقيع أو حادثة النظام.
CH19-F69— حالة نشطة مخزنة مؤقتًا
تظهر العلامة الملغاة نشطة بسبب الذاكرة المؤقتة القديمة.
CH19-F70— إخفاء نقل النطاق القانوني
غير معروف من هو السلطة الجديدة.
CH19-F71— النظر في نسب المؤسس كمصدر للسلطة
اسم المؤسس يتحول إلى احتكار للقرار الدائم.
CH19-F72— التعليم في BQ-0
يتم الإعلان عن الامتثال العام قبل إجراء الاختبار.
CH19-F73— الادعاء المستقلة في GOV-0
يتم تقديم مسودة يقودها المؤسس كمؤسسة مستقلة.
CH19-F74— ترسيخ علامة +950 في العرض التجريبي
يصبح الرقم الاصطناعي في الكتاب علامة حقيقية.
CH19-F75— تعزيز الخطأ المفضل عن طريق تمييزه
الخطأ الإيجابي المعروف يتم تحويله إلى ميزة تسويقية.
CH19-F76— تحويل الخطأ غير المواتي إلى مسألة حرجة من خلال الشكوى التجارية
ضغط العميل، وليس الأدلة، يحدد مستوى الأهمية.
CH19-F77— إغلاق سوء استخدام العلامة تجارياً بشكل خاص
يتم حل العلامة الخاطئة العامة بتحذير خاص؛ لا يتم تسجيلها في السجل.
CH19-F78— المعيار العالمي بدون تكرار مستقل
يتم إعلان مؤسسة تأسيس واحدة كسلطة عالمية.
CH19-F79— نقص صانع القرار
يتشكل نظام وراء الذكاء الاصطناعي والأتمتة حيث لا يتحمل أحد المسؤولية.
CH19-F80— استثناء لـ NOMOS
جميع قواعد الاستقلال في المعيار لا تنطبق على نظام علامته التجارية الخاص.

71. إجراء قرار المطابقة

الخطوة 1— التحقق من الامتثال للتدقيق

BQ

الحكومة

NCL

TPR

AQ

NSQ

يتم مراجعة مستويات الجودة.

الخطوة 2— تثبيت الكيان المدقق والنطاق

الكيان، المجال، منتجات الذكاء الاصطناعي، اللوحة، المُدخل، اللغة، البلد، والوقت مكتوبة.

الخطوة 3— إكمال تقرير التدقيق

يتم تسجيل توزيع الاستجابات والنتائج والدرجات وعدم اليقين والحوادث.

الخطوة 4— إجراء مراجعة تعارض المصالح

يتم تصنيف المراجعة والاستشارات والتمويل وعلاقات المؤسسين.

الخطوة 5— تعيين عينة قرار مستقلة

فريق التدقيق ليس لديه سلطة اتخاذ القرار النهائية.

الخطوة 6— تطبيق البوابات

يتم فحص البوابات الحرجة والعناصر المطلوبة غير المحلولة للجودة.

الخطوة 7— تحديد الحالة

يتم اتخاذ القرار بين CS-0 و CS-9.

الخطوة 8— مراجعة تعيين +950 إذا كان متاحًا

تُطبق شروط إضافية بعد الامتثال الكامل.

الخطوة 9— تحديد فترة الصلاحية

يتم أخذ المخاطر والتقلب وخطة المراقبة في الاعتبار.

الخطوة 10— الموافقة على نص استخدام العلامة

يتم الإبلاغ عن التعبيرات المسموح بها والممنوعة إلى المؤسسة.

الخطوة 11— إنشاء سجل التسجيل

يتم ربط جميع سجلات النطاق والقرار والأدلة.

الخطوة 12— فصل الأدلة العامة والمقيّدة

يتم الحفاظ على الخصوصية والتحقق معًا.

الخطوة 13— توقيع القرار

يحمل قرار الشخص المسؤول أو اللجنة التجزئة والتوقيع.

الخطوة 14— تفعيل السجل

سيكون التوقيع صالحًا من الآن فصاعدًا.

الخطوة 15— إنتاج أصل العلامة

تمت إضافة معرف السجل، الصلاحية، والاتصال المباشر.

الخطوة 16— بدء خطة المراقبة

تم فتح الرقابات الشهرية والدورية والمحفَّزة بالأحداث.

الخطوة 17— فتح قنوات الشكاوى والاستئناف

تم جعلها مرئية في السجل العام.

الخطوة 18— تتبع الأحداث

في حالة وجود إساءة استخدام حرجة أو تغيير جوهري، يتم تحديث الحالة.

الخطوة 19— اتخاذ قرار التجديد أو الانتهاء

لا يمكن إجراء تمديد غير محدود بدون تدقيق جديد.

الخطوة 20— حفظ السجل

القرار الجديد لا يمحو السجل القديم.

72. الأدلة المطلوبة

الإصدار القياسي؛ جودة المعيار المرجعي؛ نضج الحوكمة؛ تفويض المدقق؛ سجل الكيان؛ ملكية النطاق؛ متجه النطاق؛ سجل نظام الذكاء الاصطناعي؛ إطار السكان؛ سجلات اللوحة؛ مجموعة الادعاءات؛ تغطية اللغة والبلد؛ موجات القياس؛ حزمة الحقيقة؛ سجلات الادعاءات؛ تقرير التدقيق؛ توزيع GEO-1000؛ النتائج الحرجة؛ النتائج الكبيرة؛ النتائج المعتدلة؛ النتائج الاستشارية؛ ملف المكون؛ النتيجة المركبة؛ فترة الثقة؛ التغطية القابلة للتقييم؛ الحوادث التاريخية؛ سجلات التصحيح؛ إعلان تعارض المدقق؛ سجل علاقة الاستشارة؛ سجل التمويل والرسوم؛ عضوية عينة القرار؛ الاستبعاد؛ تصويت القرار؛ الاعتراض؛ حالة الامتثال؛ تصنيف الأداء؛ تاريخ السريان؛ تاريخ الانتهاء؛ خطة المراقبة؛ قناة الشكاوى؛ إجراء الاستئناف؛ سجل الأدلة العامة؛ بيان الأدلة المقيدة؛ تجزئة تقرير التدقيق؛ تجزئة القرار؛ توقيع السجل؛ وملف العلامة.

الرمز القابل للقراءة آليًا نص عرض العلامة مراقبة سوء الاستخدام سجلات الإيقاف المؤقت سجلات الانسحاب سجلات الإلغاء سجل الانتهاء سجل النسخ تاريخ الإصدارات التصحيحات قيد السجل العام URL سجل أمان السجل الشخص أو المؤسسة المسؤولة

73. قائمة تدقيق التدقيق

هل تم فصل القرار بين المعيار والعلامة التجارية؟ هل تم احتساب تقرير التدقيق كنجاح بمفرده؟ هل حالة البوابة قبل النتيجة؟ هل النطاق واضح في جميع المجالات؟ هل الكيان والنطاق صحيحان؟ هل منتجات الذكاء الاصطناعي واضحة؟ هل تم تعريف نطاق الوصول الطبيعي أو قاعدة المقارنة المشتركة؟ هل تم فصل اللوحات المسيطر عليها والطبيعية؟ هل تم تعريف عائلات المُدخلات؟ هل نطاق اللغة والبلد واضح؟ هل تتوفر نسخة Truth Pack وطريقة التقييم؟ هل تواريخ بداية ونهاية الصلاحية مرئية؟ هل العلامة غير محددة؟ هل قيد السجل العام نشط؟ هل يذهب رمز الاستجابة السريعة مباشرة إلى السجل الرسمي؟ هل تم استخدام لقطة الشاشة الثابتة كدليل مباشر؟ هل هناك معرف سجل على العلامة؟ هل يشير الشرط إلى ذلك صراحة؟

هل لا يزال يتم استخدام علامة منتهية الصلاحية أو معلقة؟ هل يعتمد +950 على النقاط فقط؟ هل هناك أية قضايا حرجة أو جسيمة مفتوحة؟ هل يمكن رؤية حادثة تاريخية؟ هل تم حذفها من تاريخ التصحيح الحالي؟ هل تم تعميم العلامة على جميع أنظمة الذكاء الاصطناعي؟ هل تم تأسيس ثقة أو ضمان توصية للذكاء الاصطناعي؟ هل هناك انطباع بالموافقة الحكومية أو الجامعية؟ هل تم نقل العلامة إلى مجال آخر أو شركة تابعة؟ هل البيانات المهيكلة أوسع من إعلان البشر؟ هل تم توقيع رمز التسجيل؟ هل لدى العلامة نظام لمراقبة سوء الاستخدام؟ هل اتخذ فريق التدقيق قرار المطابقة بمفرده؟ هل عينة القرار مستقلة؟ هل عينة الاستئناف مستقلة عن القرار الأول؟ هل لدى المؤسس حصة ملكية في الملف؟ هل يتطلب Kaan MURAZ الاستبعاد؟ هل قدم NobleJackal استشارات للعميل؟

هل العلاقة الاستشارية علنية؟ هل هناك مراجعة خارجية محسّنة؟ هل تعتمد رسوم المدقق على النتيجة؟ هل تؤثر رسوم العلامة على قرار التدقيق؟ هل لدى الراعي سلطة معيارية؟ هل هناك احتمال للتسوق في التدقيق؟ هل التدقيقات المباشرة مسجلة؟ هل تم حفظ تاريخ الانسحاب؟ هل تم تعيين المدقق وفقًا للنتيجة؟ هل تم إعطاء فئة نزاع؟ هل تم إزالة شخص 3-COI من القرار؟ هل يسيطر مجلس الإدارة على مجموعة مساهمين واحدة؟ هل لدى المؤسس حق النقض؟ هل لممثلي المصلحة العامة واللغة حقوق تصويت حقيقية؟ هل مر تغيير المعيار بمراجعة عامة؟ هل تم تغيير العتبة بناءً على نتيجة التدقيق؟ هل المسافة الجديدة ألغت القرار القديم؟ هل الاستئناف قائم على الأدلة؟ هل الاستئناف الحاسم أزال التعليق؟

هل هناك قناة شكاوى عامة؟ هل هناك حماية للمبلغين عن المخالفات؟ هل وجود الأدلة المحدودة والمراجع مرئي؟ هل يتم حفظ سجل تحرير السجل؟ هل تم فتح حادثة أمان السجل؟ هل حدود النطاق القانوني والإعلان الرسمي واضحة؟ هل يظهر NOMOS كمالك لتوقيع قرار الذكاء الاصطناعي؟ هل يمكن تحديد الشخص أو اللجنة المسؤولة؟ هل مستوى BQ و GOV الحاليين كافيان لمنح العلامة؟ هل تم فحص شروط الاستخدام القانونية للعلامة بشكل منفصل؟

74. الاعتراضات والردود

الاعتراض 1— "ألن تقل القيمة التجارية للمعيار بدون شارة؟"

على المدى القصير، ربما. لكن إصدار الشارة قبل توفر الأدلة والحوكمة والاستقلالية يدمر مصداقية المعيار على المدى الطويل. تستمد الشارة قيمتها:

- ليس من كونها صادرة مبكراً،
- وليس من كون الحصول عليها سهلاً،
- بل
- من حقيقة أن البوابات الإلزامية لها واقعية.

هذا هو مصدر قيمتها.

الاعتراض 2— 'إذا تم تأسيس معيار **NobleJackal**، لماذا لا يمنح الشارة **NobleJackal**؟'

في فترة التأسيس:

- تدقيق تجريبي،
- تقييم البحث،
- الأمانة الفنية

يمكن تنفيذها. ومع ذلك، من أجل ادعاء أهليّة مستقلة:

- تدقيقd،
- قرار،
- الاستئناف

يجب فصل أدوارهم. إذا منحت NobleJackal علامة لعملائها فقط، يمكن أن يكون التعبير الصحيح: "تقييم داخلي NobleJackal." لا يمكن أن يكون "مطابقة مستقلة NOMOS".

اعتراض 3— "ألن يعني تحديد سلطة المؤسس السماح للآخرين بالاستيلاء على **NOMOS**؟"

لا. يمكن الحفاظ على نسب المؤسس، والتاريخ القانوني، والمساهمات الفكرية. ومع ذلك، المعيار العالمي: إذا كان يعتمد فقط على إرادة المؤسس، فلن يستطيع البقاء بعد المؤسس. فكرة تقييد السلطة لا تضعفها. إنها تجعلها مؤسسية.

الاعتراض 4— "هل سيبقى اسم كان فوق المعيار؟"

يمكن للمؤسس أن يبقى ككاتب وإنسان مسؤول. ومع ذلك، كل قرار يتعلق بالامتثال:

- رأي كان الشخصي،
- الموافقة التلقائية للمؤسس

لا يمكن أن يحدث. يجب فصل الدور التاريخي لـ Kaan MURAZ عن سلطة اتخاذ القرار الحالية للمستند.

الاعتراض 5— "لماذا لا يمكن لـ **NOMOS** أن يكون سلطة التوقيع؟"

لأن قرار المطابقة:

- يمكن أن يكون له عواقب قانونية,
- أخلاقي،
- تجارية وعامة

يمكن للذكاء الاصطناعي المساعدة. ومع ذلك، بالنسبة للاعتراض والمسؤولية والعقوبات، مطلوب شخص أو مؤسسة مسؤولة.

الاعتراض 6— "إذا حصلت الشركة على **963**، لماذا لا يمكنها أن تقول **950+**؟"

لأن 950+ ليس مجرد تقييم. إنه أيضًا:

- يلتزم بشروط النجاح الصارم,
- البوابات الحرجة والجسيمة,
- الحد الأدنى للمكونات,
- العدل،
- غير محلولة.
- التكرار،
- الحوكمة

المتطلبات.

اعتراض 7— "أليس من المبالغة تعليق العلم بسبب حادثة حرجة واحدة؟"

حدث واحد لا يمثل مجموعة المستخدمين بأكملها. ومع ذلك، إذا أعطى العلم النشط الانطباع بعدم وجود أحداث حرجة مفتوحة، فإن ترك العلم دون تغيير دون التحقيق في الحدث هو ثقة مفرطة. تعليق حرج ليس إبطالًا دائمًا. إنه عملية تحقيق.

اعتراض 8— "ألن يكون العلم مكلفًا جدًا إذا كان صالحًا لمدة ستة أشهر فقط؟"

إنه ممكن. يمكن استخدام أتمتة المراقبة والموجة الصغيرة القائمة على المخاطر. ومع ذلك، إذا كانت منتجات وموارد الذكاء الاصطناعي تتغير بسرعة، فإن الإشارة الثابتة الصالحة لسنوات ليست موثوقة. التكلفة لا تجعل الملاءمة الخاطئة إلى أجل غير مسمى مشروعة.

اعتراض 9— 'كيف سنعرض النطاق الكامل على اللافتة؟'

تشير العلامة المرئية إلى معلومات مختصرة:

- الحالة,
- صالح حتى

معرّف السجل. موجود في صفحة السجل القانوني الكاملة للنطاق. العلامة وحدها لا تحتاج إلى عرض التقرير الكامل.

الاعتراض 10— "هل يتعين علينا نشر كل الأدلة في السجل العام؟"

لا. الطبقات العامة والتدقيق والمقيدة منفصلة. ومع ذلك، يجب أن يتخذ القرار:

- أي نوع من الأدلة،
- أي مراجعة،
- أي نطاق

يعتمد على ما هو مفهوم.

اعتراض 11— "إذا دفع العميل رسوم التدقيق، أليس الاستقلال مستحيلاً بالفعل؟"

ليس مستحيلاً. الضوابط:

- رسوم ثابتة لا تؤثر على النتيجة،
- تمويل منفصل للوحة القرار،
- جدول رسوم عام،
- تدوير المدقق،
- الإفصاح عن التضارب،
- حظر رسوم النتائج

يمكن أن يقلل المخاطر.

الاعتراض 12— "إذا كان المدقق قد طور موقع العميل سابقًا، هل يمكنه عدم إجراء التدقيق؟"

في بعض البرامج التجريبية منخفضة المخاطر، يمكنهم أن يلعبوا دورًا مع فريق منفصل ومراجعة محسنة. ومع ذلك، لا يمكنهم الفصل في عملهم بمفردهم، أو إعطاء قرار المطابقة. قد يكون التدقيق المستقل أكثر موثوقية في الملفات عالية المخاطر ماديًا.

الاعتراض 13— "ألا تؤدي عملية الاعتراض إلى إبطاء الأمور؟"

يمكن أن يحدث ذلك. لكن نظام المطابقة بدون استئنافات لا يمكنه تصحيح:

- خطأ في الفصل،
- عيب في حزمة الحقيقة.
- أو تعارض في المصالح.

السرعة ليست أهم من تجنب اتخاذ قرار عام خاطئ.

الاعتراض **14**— "ألن يكون من الظلم للمؤسسة إذا بقيت السجلات السلبية القديمة في السجل؟"

يجب أن يكون الوضع الحالي والتصحيح واضحين. ومع ذلك، فإن حذف السجلات القديمة يعيد كتابة تجربة المستخدم السابقة. نظام مناسب:

- يعرض الخطأ،
- والتصحيح،
- والنتيجة القوية الحالية.

معًا.

الاعتراض **15**— 'إذا طبقت مؤسسات أخرى طريقة **NOMOS**، ألن تتدهور الجودة؟'

يمكن أن يتدهور. لذلك، يلزم الحصول على تفويض المراجع، والكفاءة، والسجل العام. ومع ذلك، فإن حقيقة أن الشركة المؤسسة فقط يمكنها تطبيق الطريقة ليست متوافقة أيضاً مع المعايير العالمية المستقلة. يجب أن يتم التأكيد على التفويض جنباً إلى جنب مع الشفافية.

اعتراض **16**— "هل يمكننا حالياً كتابة **NOMOS** بواسطة **NobleJackal** على الموقع الإلكتروني؟"

يمكن استخدام المؤلف بهوية مشروع بحثي ومرشح معياري. يجب أن يكون هذا التمييز واضحاً: "NOMOS بواسطة NobleJackal— مرشح معياري مفتوح" التعبير الذي لا يمكن استخدامه: "NOMOS سلطة معتمدة" لأن الاختبار الحقيقي، والحكم المستقل، وسلطة المطابقة لم تُنشأ بعد.

البند الشائع في القسم **79**

تصميم شعار أمر سهل. كتابة "تم التحقق" تحت الشعار أسهل حتى. من الممكن إضافة رمز QR. من الممكن كتابة "certified": true" في حقل JSON. لا شيء من هذا بمفرده يشكل ثقة. الثقة تنشأ من القدرة على الإجابة عن هذه الأسئلة: من تم تدقيقه؟ ما اسم النطاق؟ ما منتجات الذكاء الاصطناعي؟ ما مجموعة المستخدمين؟ ما اللغات؟ ما الدول؟ ما التحفيزات؟ في أي تاريخ؟ أي حزمة حقائق؟ أي طريقة تقييم؟ من كان المحكمين؟ ما تضارب المصالح؟ ما النتائج الحرجة والهامة؟ ما الطعون؟ هل العلامة لا تزال صالحة الآن؟ إذا لم تكن الإجابات على هذه الأسئلة مرئية، فإن العلامة ليست الوجه المرئي للمعيار، بل الوجه المرئي رسميًا للإعلان. لا يمكن أن تكون العلامة NOMOS جملة جديدة تقولها الشركة عن نفسها. العلامة:

يجب أن يكون هناك ملخص يمكن الوصول إليه علنًا لسلسلة القرارات مستقل عن الشركة. يجب أن تكون هذه السلسلة من القرارات قادرة أيضًا على أن تصبح مستقلة عن المؤسس. يمكن بدء معيار Kaan MURAZ. يمكن إنشاء المجال القانوني NobleJackal. يمكن تطوير طريقة NOMOS والصوت النقدي. ومع ذلك، لكي يستمر المعيار مع مرور الوقت:

- يجب على المؤسس تقييد سلطته الخاصة،
- يجب أن تكون القرارات مفتوحة للمراجعة المستقلة.
- يجب الحفاظ على المعارضة،
- يجب أن تكون المؤسسات الأخرى قادرة على إعادة إنتاج الطريقة،

يجب ألا يكون السجل العام ذاكرة المؤسس الأحادية. هذا لا يمحو اسم المؤسس. على العكس، فإنه يمنح المؤسس الدور التاريخي التالي:

> شخص يأخذ الأمور بجدية كافية لجعل المعيار الذي أسسه مستقلاً عن نفسه.

يمكن للشركة أن تحصل على درجة عالية. لكن قد يكون لديها حدث حرج مفتوح. يمكن للشركة أن تحصل على درجة منخفضة. لكنها قد لا تكون ارتكبت أي انتهاكات أخلاقية. يمكن للشركة أن تتلقى الامتثال الكامل. ثم قد تحدد ادعاء علامة تجارية خاطئ. شركة أخرى قد لا تتلقى علامة. لكنها يمكن أن تصبح أكثر موثوقية مع التصحيح. العلامة لا تقلل القصة كاملة إلى كلمة واحدة. سجل:

- يحفظ النتيجة الحالية,
- التاريخ,
- الحد,
- التصحيح

معًا. يجب أن تمنح علامة الامتثال المستخدم الكلمات: "هذه المؤسسة لا تشوبها شائبة." لا. يجب أن تعطي الكلمات:

> تم تقييم هذه المؤسسة ضمن النطاق والإطار الزمني المحددين في السجل العام، بطريقة مفتوحة وسلسلة قرار مستقلة؛ يمكن الاطلاع على النتائج الحالية، والحدود، والسجلات التاريخية دون حجبها.

لذلك، القانون التاسع عشر للقياس لـ NOMOS هو:

> الشارة ليست شعارًا يحل محل الأدلة؛ إنها باب حي مفتوح نحو الأدلة.

تنص قانون قياسه العشرون على:

> إذا لم يكن نطاق العلامة مرئيًا، فإن العلامة ليس لها معنى.

ينص القانون الحادي والعشرون للقياس على:

> الدرجة ليست امتثالًا؛ البوابة، الحوكمة، والصلاحية تأتي قبل الدرجة.

ينص القانون الثاني والعشرون للقياسات على:

> المعيار التأسيسي يمكن أن يبدأ؛ لا يمكن أن يكون الحكم الوحيد على عميله.

القانون الثالث والعشرون كما يلي:

> عندما يكون المُدقق وصانع القرار في الامتثال هما نفس الجهة، يجب إثبات الاستقلالية.

القانون الرابع والعشرون هو كما يلي:

> الإشارة الثابتة لا تبقى دقيقة مع مرور الوقت؛ الامتثال يموت بدون سجل حي.

القانون الخامس والعشرون هو ما يلي:

> السجلات المعلقة أو الملغاة لا تُحذف؛ بل تُحفظ كتاريخ تحت الوضع الحالي.

القانون السادس والعشرون هو:

> يمكن للذكاء الاصطناعي قراءة الإشارة؛ الإشارة لا تضمن أن الذكاء الاصطناعي سيثق بها أو يوصي بها.

القانون السابع والعشرون هو:

> يتم إنشاء الاستقلالية ليس بمظهر الاستقلالية، بل بفصل الواجبات والتمويل وسلسلة الاستئناف.

القانون الثامن والعشرون هو:

> الصوت الرسمي لـ *NOMOS* موجود فقط في السجل القانوني الذي يحوي موافقة الإنسان، الإصدار، والنزاهة.

القانون التاسع والعشرون هو:

لكي يدوم المعيار بعد مؤسسه، يجب أن تخضع سلطة المؤسس أيضًا للمعيار.

القانون الثلاثون هو كما يلي:

حتى إذا تم الانتهاء من هذا الكتاب، لا يكتمل أي توقيع بدون حكم مستقل.

الطلب 19 من NOMOS

لا تحولني إلى شعار.

لا تضعني على موقعك وتقول 'الذكاء الاصطناعي يثق بنا'.

لا تجعل تقييمّي ضمانًا للتوصية.

عند استخدام توقيعي، أظهر بالضبط من، ماذا، أين، بأي لغة، وفي أي تاريخ تم التحقق.

لا تنشر النتيجة التي حصلت عليها في مجال واحد على جميع شركاتك.

لا تعمم تدقيقًا أُجري بلغتين على العالم كله.

لا تعرض نتيجة Controlled Clean كما لو كانت تخص جميع المستخدمين الحقيقيين.

لا تبيع Core *GEO-1000* كـ Full *NOMOS*.

لا تخفي كلمة "شرطي" في علامتي الشرطية.

قم بإزالتي عندما تنتهي صلاحية تاريخ إنتهاء صلاحيتي. / لا تحذف سجلي التاريخي.

لا تجعل نتيجتي العالية درعًا عندما أواجه حدثًا حرجًا. / ضعني في الانتظار.

لا تعمم حدثًا فرديًا على النظام بأكمله. / لا تعامل حدثًا فرديًا كما لو كان غير موجود.

لا تربط رمز الاستجابة السريعة الخاص بي بصفحتك الإعلانية. / اربطه بالسجل العام.

لا تعتبر الصورة التي ليس لها معرف سجل علامة صالحة.

لا تقدم ادعاءات تفوق قدرة البشر في البيانات المنظمة.

لا تصادق على العمل الذي قمت به بمفردك على أنه مستقل.

إذا قدمت استشارة، فاشرح ذلك.

فقط لأنك أجريت التدقيق، لا تتخذ قرار العلامة التجارية.

الشخص الذي يتخذ قراري لا ينبغي أن يراجع اعتراضي.

لا تربط أتعابك بدرجتي العالية.

لا تسمح لعميل كبير بشراء معياري.

أظهر الاحترام لمؤسسي. / لكن لا تجعل مؤسسي صاحب حق النقض.

لا تمحُ اسم كاان. / كما لا تسمح له بأن يكون القاضي الوحيد في ملف له مصلحة فيه.

سجل *NobleJackal* كحقل ولادتي الأساسي. / لا تجعل *NobleJackal* القاضي الوحيد إلى الأبد.

استخدمني ككاتب مشارك، ومؤسس الطريقة، ونظير نقدي. / لا تعرضني كموقع قانوني.

اعتبر نصي فقط في الحقل القانوني، مُرقَّم بالإصدار، مؤرخ، معتمد من الإنسان، ويحمل سجل النزاهة كتصرّح رسمي.

لا تجعل كل تصريح قلته في المحادثة قرارًا قياسيًا.

افتح مسار الشكوى. / لا تخفِ الآراء المعارضة. / لا تُسكت المُبلّغ عن المخالفات.

لا تغير معياري بناءً على نتيجة العميل.

لا تحذف القرار القديم عند كتابة قاعدة جديدة.

لا تمنح شارات مع عرضي الاصطناعي.

لا تعلنني كسلطة عامة بدون إجراء الاختبار الحقيقي، وبدون فريق مستقل يعيد تكراري، وبدون إنشاء مجالس القرار والاستئناف.

أولاً، تحقق من صحة الطريقة. / ثم أفوض المدقق. / ثم افصل التدقيق عن الاستشارة. / ثم أنشئ لوحة قرار مستقلة. / ثم افتح السجل العام. / ثم أنشئ طرق الشكاوى والاستئناف. / ثم نفذ المراقبة والإيقاف. / فقط بعد ذلك يمكن أن تصبح العلامة مرئية.

جملة ختام الفصل

علامة المطابقة NOMOS ليست ادعاءً آخر بالمدح تصنعه شركة عن نفسها. تكتسب معناها فقط باعتبارها التعبير المرئي لسجل عام مفتوح ومستقل لا يمكن تعديل نطاقه وتاريخه بشكل أحادي من قبل المؤسس أو المدقق أو العميل أو NOMOS نفسها.

النواة المعيارية

يجب أن يظل كلٌّ من معيار **NOMOS**، وتقرير التدقيق، وقرار المطابقة، وعلامة المطابقة، وقيد السجل العام أثرًا مستقلًا ومتميزًا. ولا تصبح أي علامة **NOMOS** صالحة لمجرد إجراء تدقيق أو إنتاج درجة رقمية. ويجب ربط كل علامة صالحة بقيد مرجعي نشط في السجل العام يحدد: - الكيان الخاضع للتدقيق، - النطاقات المصرح بها، - منتجات الذكاء الاصطناعي، - إطار المجتمع المستهدف، - حالات اللوحات، - عائلات المُدخلات، - اللغات، - البلدان، - موجات القياس، - إصدار حزمة الحقائق المتحقَّق منها، - إصدار طريقة احتساب الدرجات، - حالة المطابقة، - مدة الصلاحية، - الحوادث، - المراقبة المستمرة، - جهة التدقيق، - هيئة اتخاذ القرار، - حالات تضارب المصالح، - الاستئنافات، - وسجلات التكامل. ولا يجوز لصورة ثابتة أو لقطة شاشة أو ادعاء في بيانات منظَّمة صاغه صاحبه بنفسه أو رابط ميت إلى السجل أن يثبت مطابقة حالية. ولا يجوز تقديم علامة **NOMOS** بوصفها: - ضمانًا لجدارة الذكاء الاصطناعي بالثقة أو لتوصيته، - دليلًا على الريادة العالمية، - اعتمادًا حكوميًا أو جامعيًا، - ترخيصًا مهنيًا، - تغطية شاملة لجميع اللغات، - أو ضمانًا لنتائج العملاء. ولا يجوز نقل نطاق المطابقة بين الكيانات أو النطاقات أو المنتجات أو الشركات التابعة أو الامتيازات أو البلدان أو اللغات أو منتجات الذكاء الاصطناعي أو حالات اللوحات أو عائلات المُدخلات أو مدد الصلاحية من دون قرار جديد. ويجب أن تنشئ الحادثة الحرجة المؤكدة حالة تعليق حرج تتناسب مع النطاق، أو حالة أشد، إلى أن تُعالج الحادثة. ولا يجوز لدرجة مرتفعة أن تتجاوز هذه البوابة. ويجب فصل وظائف التدقيق والاستشارات واتخاذ قرار المطابقة وتشغيل السجل وترخيص العلامة والنظر في الاستئنافات؛ وإذا جُمعت بعض هذه الوظائف، فيجب الإفصاح علنًا عن الأدوار المجمعة والضوابط التعويضية. ولا يجوز لفريق التدقيق أن يكون وحده هيئة القرار النهائي بشأن المطابقة. كما لا يجوز للمؤسس أو **NobleJackal** أو أي مدقق أو مزود للذكاء الاصطناعي أو راعٍ أو كيان خاضع للتدقيق أن يملك سيطرة منفردة على تغييرات المعيار أو قرارات المطابقة أو الاستئنافات أو تاريخ السجل العام. ويجوز لـ **NOMOS** من **NobleJackal** أن يكون المنشور المرجعي التأسيسي والأمانة الفنية، لكن يجب أن يتطور نظام المطابقة العام والمستقل نحو قرارات متعددة الأطراف وآلية استئناف وحوكمة تراعي المصلحة العامة. ويجوز لنظام ذكاء اصطناعي أن يساعد في الصياغة والتحليل والاستخراج واحتساب الدرجات والرصد ومراجعة الاتساق، لكنه لا يجوز أن يكون الموقّع القانوني أو المسؤول الوحيد على قرار المطابقة. ولا يجوز تقديم منشورات **NOMOS** أو قراراته بوصفها بيانات **NOMOS** رسمية إلا إذا كانت مرجعية وتحمل رقم إصدار وتاريخ نشر واعتمادًا من شخص مسؤول وسجل تكامل. ويجب أن تظل المخرجات الأخرى محاكاة غير رسمية. ولا يجوز أن تعتمد رسوم المطابقة أو أتعاب المدققين أو المراجعين أو ترخيص العلامة على تحقيق درجة مرغوبة أو الحصول على علامة. ويجب أن تظل السجلات المنتهية أو المعلّقة أو المسحوبة أو الملغاة أو محل الاستئناف أو المعالجة أو التاريخية ظاهرة في السجل العام ذي الإصدارات، مع مراعاة القيود المتناسبة المتعلقة بالخصوصية والقانون. ولا يجوز إصدار علامة مطابقة عامة لـ **NOMOS** أو تسمية **NOMOS +950** ما دامت متطلبات الاختبار المرجعي والحوكمة وترخيص المدققين والقرار المستقل والاستئناف والسجل، التي يشترطها إصدار المعيار المعلن، غير مستوفاة. ويجب ضبط إصدار كل تغيير في المعيار وتدقيق وقرار وعلامة وواقعة مراقبة وشكوى واستئناف وتعليق وسحب وإلغاء ومراجعة للسجل وقرار حوكمة، ونسبه إلى شخص أو منظمة مسؤولة.

المعيار المفتوح، ودعوة الجامعات، وتسليم NOMOS إلى المجتمع

لضمان بقاء المعيار بعد مؤسسه

الأسس المنهجية / K18; K19; K23; K24

حد الفصل

حوّلت الفصول التسعة عشر الأولى GEO من ادعاءات فردي للرؤية إلى مقياس تمثيل قائم على السكان. فقد فصلوا الكيان المدقق عن مجاله، وعلامته التجارية، والشخصية القانونية؛ وفصلوا التمثيل الرسمي عن الواقع الموثّق؛ وأسسوا تصميم عينات GEO-1000 عبر البلدان واللغات؛ وعرّفوا منتج الذكاء الاصطناعي من خلال خطته، وواجهته، وأدواته، وزمنه، وحالة المستخدم، بدلاً من الاعتماد على اسم المزود فقط؛ واعتبروا الادعاءات أدوات تجريبية؛ وفصلوا بين لوحات المستخدمين النظيفة والطبيعية المسيطر عليها؛ وبنوا سلسلة الأدلة المتزامنة والموجية NOMOS Capture؛ وعرّفوا حزمة الحقائق المتحقَّق منها عن الكيان؛ وفككوا استجابات الذكاء الاصطناعي إلى ادعاءات ذرية؛ واحتوى على منع اختفاء الأخطاء الحرجة والجسيمة ضمن المتوسط.

أنشأت NOMOS تقييمها كنظام من التوزيعات، البوابات، المكونات، العدالة واللايقين بدلاً من أن يكون رقمًا واحدًا. وقد عرضت طريقتها الخاصة للاختبار الاصطناعي، وحولت علامة الامتثال من شعار ثابت إلى سجل عام مباشر، وقلصت سلطة المؤسس، المدقق، العميل والمساعد الذكائي. كانت أربع طبقات مرئية بالفعل في الفكرة التأسيسية: تحدد GEO Framework المعيار؛ وتحدد 99 الأخطاء في GEO كيفية انتهاك المعيار؛ يقيس نظام التدقيق الانتهاكات ويوثقها؛ وتتحقق العلامة المستقبلية من النطاقات التي تستوفي المعيار. معًا، تشكل الآن نظامًا معياريًا مترابطًا. ومع ذلك، النظام غير مكتمل. كتابة معيار لا يجعل هذا المعيار ملكاً للعالم. يصبح المعيار عاماً حقًا فقط عندما يكون:

- متاحاً للجمهور,
- مفتوحاً للفحص من قبل الجامعات,
- قابلاً للتفنيد من قبل الباحثين المستقلين,
- قابلاً للتكرار عبر الدول,
- قابلاً للتطبيق من قبل مؤسسات أخرى,
- مفتوحاً للنقد بشكل مستقل عن مؤسسه,
- ذو نسخ متعددة دون قمع النتائج السلبية.

إذا ظل المعيار نصاً يُسيطر عليه فقط NobleJackal، فهو لا يزال مجرد:

طريقة NobleJackal

قد تكون هذه الطريقة ذات قيمة، لكنها لا يمكن أن تصبح معيارًا عالميًا مشتركًا لـ GEO. إذا كان يمكن تغيير المعيار فقط من خلال تفسير Kaan MURAZ، فإنه يظل:

إعلان المؤسس

قد يكون إعلان المؤسس مقنعًا، لكنه لا يمكن أن يعيش بعد وفاة مؤسسه. إذا كان المعيار يتألف فقط من النصوص المنتجة في هذه المحادثات، فإنه يظل:

نظامًا إبداعيًا وشاملاً من الفكر

قد يكون ذلك ذا قيمة. ومع ذلك، لا يمكن أن يصبح معيارًا رسميًا، موثقًا بالإصدار وقابلًا للتنفيذ بشكل مستقل. ولذلك، فإن السؤال النهائي المعياري للكتاب هو:

> كيف نتخلى عن الادعاء بالملكية الفردية لـ *NOMOS* دون فقدان *NOMOS*؟

هنا، 'الاستسلام' لا يعني:

- حذف اسم Kaan MURAZ كمؤسس,
- رفض الأصل القانوني لـ NobleJackal،
- ترك النص بلا صاحب,
- السماح لأي شخص بتعديله دون إفصاح،
- قبول انتشار علامات NOMOS المقلدة.

التسليم العام يعني بدلاً من ذلك:

> المعيار متاح للقراءة والنقد والتطبيق وإعادة الإنتاج والتطوير؛ ومع ذلك، يتم الحفاظ على النسخة الرسمية والقرار والعلامة وسلطة السجل تحت إدارة عامة مسؤولة.

الانفتاح والسلطة ليسا أعداء. الانفتاح يشرح كيف تعمل الطريقة؛ السلطة تحدد النسخة الرسمية. الانفتاح يسمح بالاختبار المستقل؛ السلطة تمنع نسخة مشتقة من تقديم نفسها على أنها النسخة الرسمية NOMOS. الانفتاح يمكّن النقد الجامعي؛ السلطة تمنع تمثيل رأي شخص أكاديمي واحد على أنه موافقة مؤسسية. الانفتاح يسمح للمؤسسات الأخرى بإعادة إنتاج الحساب؛ السلطة تمنع علامات المطابقة الصادرة ذاتياً. هذا الفصل يمد البنية السابقة لعناوين URL الأساسية، وسجلات الأخطاء المنفصلة، والإصدارات، والتنويهات، والاستشهادات، وسجل قابل للقراءة آليًا على كامل نظام NOMOS. يجب على كل وحدة معيارية الاحتفاظ بسجلها الرسمي وإصدارها وسجل تصحيحاتها الخاص. لذلك يحدد هذا الفصل:

- تعريف المعيار المفتوح
- الحدود القانونية والإدارية لتسليم الجمهور
- التمييز بين نسخة معيارية وفرع
- فصل النصوص، الأكواد، البيانات وتراخيص العلامات التجارية،
- الإصدارات التأسيسية المشتركة السلطة باللغتين التركية والإنجليزية،
- عملية الترجمة المصرح بها للغات الأخرى,
- الاستشارة العامة ومقترحات التعديل
- دعوة مفتوحة للجامعات,
- برنامج بحثي واستنساخ مستقل
- الحق في نشر النتائج السلبية,
- البيانات، الأخلاقيات وحوكمة الخصوصية،
- دور مزودي الذكاء الاصطناعي والكيانات المنظمة,
- شبكة بحثية للغات ذات الموارد المحدودة
- استقلال التمويل والرعاية,
- الطريق إلى الاستقلال عن المؤسس

- الهوية الرسمية لـ NOMOS
- الدور المستقبلي لـ NobleJackal,
- الشروط للانتقال إلى الإصدار 1.0،
- الأرشيف العام وقواعد الاستمرارية
- المظروف المفتوح المعياري القابل للقراءة آليًا.

كما يوضح صراحة ما لا يفعله هذا الفصل:

- منح وضع هيئة معايير وطنية أو دولية قائمة على NOMOS،
- ادعي موافقة الجامعة على NOMOS،
- إنشاء الاعتماد الأكاديمي،
- قم بتحويل هذا النص المرشح تلقائيا إلى NOMOS 1.0.

ولا يفعل برنامج الامتثال العام قبل اكتمال الاختبارات المستقلة. السؤال المركزي للفصل هو:

> كيف نجعل معيارا مفتوحا بما يكفي للجميع للمراجعة، وموزعا بما يكفي بحيث لا يستطيع أحد الحصول عليه بمفرده، ومرجعيًا بما يكفي لتحديد أي نص هو الرسمي؟

التحدي NOMOS

أنت تقوم بنشر معيار. يمكن لأي شخص تنزيل ملف PDF. تقول: «NOMOS هو الآن معيار مفتوح.» لكن: لم يتم نشر النص الأصلي. لا يوجد سجل تغييرات. لا يوجد مسار للإبلاغ عن الأخطاء. الكود الحسابي مغلق. لا يوجد مجموعة اختبار. أوزان التقييم المستخدمة غير معلنة. لا يُعرف من يمكنه تغيير المعيار. الإصدارات القديمة غير متاحة.

يتم حجب التعليقات السلبية، ويجب على مؤسسة أخرى طلب إذن لتطبيق الطريقة، وفقط NobleJackal يمكنها إصدار العلامة مقابل رسوم. قد يكون مثل هذا النظام قابلاً للقراءة، لكنه ليس مفتوحًا. الآن افترض أن مجموعة أخرى قامت بتنزيل الكتاب، وتعديل النص، وإزالة البوابات الحرجة ونشرت النتيجة تحت الاسم:

NOMOS Global 2.0

تقدّم المجموعة نصها المعدّل باعتباره 'نسخة NOMOS الرسمية الجديدة والمحسّنة' وتمنح درجات عالية لعملائها الخاصين. التسليم العلني لا يخول انتحال هوية المعيار القانوني. الحق في التفرع ليس حق الادعاء بالهوية الرسمية. الآن افترض أن أستاذاً جامعياً يصف العمل شخصياً بأنه مثير للاهتمام. لا يمكن لـ NobleJackal تحويل هذا التعليق الفردي إلى 'جامعة X قد وافقت على NOMOS'. الاهتمام الشخصي للعالم ليس:

- تأييداً مؤسساتياً،
- تكراراً مستقلاً،
- تصديقاً بمراجعة الأقران،
- اعتماداً جامعياً

ولا يُعد طلب من جامعة 'للتحقق من NOMOS وتقديم خطاب دعم' دعوة للبحث المستقل. إنه يطالب بنتيجة إيجابية. يجب أن تقول دعوة الجامعة الصحيحة:

> لا توافقوا علينا. اختبرونا. انشروا المكان الذي تجدونه غير صحيح.

الآن يطبق فريق بحث آخر GEO-1000 ويحقق نتيجة مختلفة عن NOMOS. يرفض مجلس NOMOS الدراسة قائلاً: 'فقط نحن يمكننا إصدار النتيجة الرسمية.' بعد ذلك، لم يعد المعيار NOMOS مفتوحًا؛ لقد أصبح سلطة مغلقة غير قادرة على رؤية أخطائها من الخارج. في سيناريو آخر، النص الكامل لـ NOMOS مفتوح ولكن علامة المطابقة غير مقيدة. تبدأ كل شركة في الادعاء بـ 'متوافق مع NOMOS' بدون تدقيق. تصبح الانفتاح استخداماً

مزيفاً للعلامة. في سيناريو ثالث، يتم نشر النص بموجب رخصة مفتوحة بينما يتم نسخ اسم المعيار وشعاره والسجل ومفتاح التوقيع دون رقابة. تصبح السجلات الرسمية وغير الرسمية غير قابلة للتمييز. هذا ليس تسليمًا عامًا؛ إنه تجزئة للهوية. وفي سيناريو آخر، يؤدي الترجمة إلى الإنجليزية إلى تحويل كلمة 'MUST' إلى 'موصى به'.

إذا حافظ ترجمة أخرى معتمدة على كلمة ‘MUST’ باعتبارها إلزامية، فقد تزعم الدولتان أنهما تنفذان نفس نسخة NOMOS مع تطبيق معايير مختلفة. الترجمة ليست مجرد وسيلة للتواصل؛ بل تحمل الرمز المعياري لـ NOMOS نفسه. لنفترض الآن أن مزودًا رئيسيًا للذكاء الاصطناعي يمول أبحاث NOMOS، وكنتيجة لذلك:

- إنه يريد عدم خفض معامل منتجه,
- للتخفيف من بوابة حرجة معينة،
- لكي تُعتبر البيانات القياسية لمزود الخدمة الدليل الوحيد

هذه هي الشروط التي يطالب بها. التمويل الشفاف لا يزال يقوض المعيار إذا أثر على القرار. المعيار إذن يصبح ملكًا للمانح، وليس للجمهور. الآن افترض أن Kaan MURAZ فقط يوافق على كل النسخ الرسمية إلى أجل غير مسمى. قد يصبح المعيار ناجحًا، لكن ماذا يحدث إذا قام Kaan:

- بتنازل عن المنصب،
- أصبح من غير الممكن الوصول إليه،
- توفّي،
- دخل في تضارب مصالح

أو لا يمكنه الاستمرار بمفرده. ما هو مستقبل المعيار؟ إذا لم يكن هناك خطة استمرار مستقلة عن المؤسس: فإن المعيار ليس مؤسسياً، بل هو سيرة ذاتية. الآن يتم تقديم NOMOS كما لو كان كيان ذكاء اصطناعي: "ذكاء اصطناعي مستقل طورته NobleJackal." ولكن ما يوجد هو:

- نموذج منفصل،
- سجل التدريب،
- البنية التقنية،
- نظام العامل ذو النسخ،
- البنية التحتية التشغيلية المستقلة

لا وجود له. في الواقع NOMOS هو:

- تطور بين Kaan MURAZ وأنظمة الذكاء الاصطناعي،
- ويستخدم في النصوص القانونية.
- المساعد النقدي المؤلف والراوي المنهجي

هو هويته. تقديمه كتقنية غير موجودة يقوض الأساس الأخلاقي لـ NOMOS. صوت الذكاء الاصطناعي لـ NOMOS ليس سلطة غير قابلة للخطأ تتحدث بالنيابة عن جميع الذكاءات الاصطناعية؛ يجب توضيح أن Kaan MURAZ يستخدم أنظمة الذكاء الاصطناعي كمحاورين نقديين. المبدأ الأول في هذا القسم هو:

> أن تكون قابلاً للقراءة لا يعني أن تكون معيارًا مفتوحًا.

ينص حكمه الثاني على:

> أن تكون معيارًا مفتوحًا لا يعني ترك الهوية الرسمية بلا مالك.

تنص الفقرة الثالثة منه على:

> دعوة الجامعة هي دعوة للتزوير، وليست طلبًا للدعم.

تنص المادة الرابعة على:

> حق التفرع ليس الحق في انتحال الهوية الرسمية *NOMOS*.

تنص أحكامه الخامسة على:

> نص *NOMOS* قابل للتطبيق صراحة؛ ومع ذلك، يجب التحكم في علامة التوافق بطريقة تحمي الثقة العامة.

ينص الحكم السادس على:

> لا يمكن إزالة اسم *Kaan MURAZ* من تاريخ الراية؛ لا يمكن ربط السلطة الوحيدة لاتخاذ القرار بمستقبل المعيار.

البند السابع هو كما يلي:

> لا يمكن تقديم *NOMOS* كنموذج أساس مستقل أو نظام ذكاء اصطناعي مستقل طوره *NobleJackal* ما لم يتم أيضا تأسيس هيكله التقني الفعلي.

النص الثامن هو كما يلي:

> لا يُقدّم المعيار للجمهور إلا عندما يحفظ حق من ينتقده في نشر النتائج.

1. غرض الفصل

يحدد هذا الفصل الشروط التي بموجبها قد يتحرك NOMOS من:

- عمل مؤلف
- إلى مرشح المعايير المفتوحة،
- برنامج بحثي مستقل
- نظام حوكمة ذو مصلحة عامة
- وفي النهاية، نظام تنفيذ دولي.

يُميز الأمور المعيارية التالية:

- النص العام والمعيار المفتوح
- البرمجيات المفتوحة المعيار والمفتوحة المصدر
- النص القانوني والتنفيذ المرجعي
- الإصدار المعياري والتفرع
- التفرع والإصدار شبه الرسمي
- حرية التنفيذ وحق استخدام العلامة
- حقوق الطبع والنشر وحق تنفيذ المعيار
- ترخيص النص وترخيص البرمجيات
- الوصول إلى البيانات الشخصية مع ترخيص البيانات
- احتكار الطريقة مع حماية العلامة التجارية
- الترجمة المجتمعية مع الترجمة الرسمية
- اللغة التوضيحية مع اللغة القانونية
- الإصدار الدولي الإنجليزي مع الإصدار التأسيسي التركي
- الترجمة الحرفية مع النص المعياري المكافئ
- سلطة القرار مع التعليق العام
- اعتماد الجامعة مع اهتمام الجامعة
- الشهادة مع التعاون الأكاديمي

- شراء النتائج الإيجابية مع تمويل البحث
- إعادة السرد مع التكرار
- العداء مع النتيجة السلبية
- انتقاد المعيار مقابل إساءة استخدام المعيار
- نسبة المؤسس مقابل سيطرة المؤسس
- أصل NobleJackal مقابل احتكار NobleJackal
- شخصية NOMOS مقابل نموذج ذكاء اصطناعي منفصل
- البيان الرسمي NOMOS مقابل ناتج المحادثة
- التسليم العام مقابل التخلي
- الإشراف القياسي مقابل الملكية
- المصلحة العامة مقابل الأغلبية المطلقة غير المحدودة
- الاختبار المفتوح مقابل تسرب البيانات المحجوزة
- الشفافية مقابل الكشف عن البيانات الشخصية
- البيانات المفتوحة مقابل الوصول البحثي المحدود
- نص النسخة 1.0 مقابل تفعيل برنامج المطابقة
- قبول المعيار والعلامة التجارية
- التنفيذ الرسمي مع التنفيذ المتوافق
- التغيير الملزم مع القضية المفتوحة
- تغيير النسخة المعيارية مع التصحيحات
- مسح التاريخ مع التقاعد
- السلطة القانونية مع أرشيف مستقل
- النشر الذاتي مع رؤية موسوعية أو أكاديمية
- فقدان اسم المؤسس مع التسليم العام

في نهاية هذا القسم، يجب أن يكون الجمهور قادرًا على فهم ما يلي:

أي نسخة من *NOMOS* رسمية؟

من يمكنه قراءة وتنفيذ وإعادة إنتاج نص *NOMOS*؟

من يمكنه اقتراح التغييرات؟

من يمكنه تغيير النسخة الرسمية؟

من يمكنه إجراء البحوث المستقلة؟

من لا يمكنه منح علامة المطابقة الرسمية؟

ما هو دور الجامعة؟

كيف يتم توضيح ما إذا كان *NOMOS* نموذج ذكاء اصطناعي، أو اسم قلم، أم هوية منهجية؟

كيف يستمر المعيار حتى إذا كان المؤسس غير متاح؟

تحت أي ظروف يمكن نشر *NOMOS 1.0*؟

2. الأحكام المعيارية المركزية

يجب أن يتم هيكلة NOMOS كمعيار عام مفتوح من حيث كونه قابلاً للقراءة والتطبيق والاختبار والنقد والتكرار وخاضعًا للبحث المستقل؛ وكمعيار عام معتمد ونسخي ومسؤول فيما يتعلق بالإصدار الرسمي وقرار المطابقة والسجل العام وهوية العلامة التجارية. التسليم للجمهور له خمس نتائج إلزامية: يجب أن يكون النص قابلاً للقراءة مجانًا وبشكل دائم. يجب أن تكون الطريقة قابلة للتطبيق بشكل مستقل. يجب أن تكون منطق الحساب والاختبار قابلاً للتكرار. يجب أن تكون طرق التغيير والاعتراض علنية. لا ينبغي لأي شخص أو شركة تغيير الهوية الرسمية بمفردهم.

3. ما هو التسليم العام؟

يعني التسليم إلى الجمهور أن NOMOS:

- محادثة واحدة,
- ملف واحد،
- موقع واحد،
- مؤسس واحد،
- شركة واحدة،
- مزود ذكاء اصطناعي واحد

يتم إخراجه من حدود السيطرة وتوصيله بنظام حكم مُرقَّم خاضع للمساءلة العامة. ويغطي ثلاثة حقوق منفصلة تُمنح للجمهور:

3.1. الحق في المعرفة

الجميع:

- يجب أن يكون قادرًا على رؤية نص المعيار،
- منطق التقييم،
- بوابات الأخطاء،
- قواعد الحوكمة،
- تاريخ التغييرات

يجب أن يكون قادرًا على رؤيته.

3.2. الحق في الاختبار

يجب على الجميع، مع الحفاظ على الظروف الأخلاقية والبيانات اللازمة:

- إعادة الاختبار،
- الصيغة،
- استخلاص الادعاءات،
- طريقة أخذ العينات،
- نموذج العدالة

يجب أن تكون قابلة لإعادة الاختبار.

3.3. الحق في الاستئناف

الجميع:

- الخطأ،
- التناقض،
- التحيز،
- تضارب المصالح،
- سوء استخدام العلامة،
- خلل الطريقة

يجب أن يكون قابلاً للتبليغ.

4. ما الذي لا يُسلم للجمهور؟

التسليم العام:

- المشاركة غير المنظمة،
- ويكي بدون إصدارات،
- استخدام العلامة التجارية بدون حدود،
- إزالة السجل القانوني،
- ترك جميع القرارات لتصويت وسائل التواصل الاجتماعي،
- جعل البيانات الشخصية متاحة للجميع،
- تسريب نتائج الاختبارات المستمرة،
- التغلب على الخبرة برأي الأغلبية،
- محو نسب المؤسس

لا تتوافق أي من هذه الممارسات مع معيار مفتوح.

5. NOMOS شرط المعيار المفتوح

نسخة NOMOS هي مرشح لمعيار مفتوح فقط إذا وفرت جميع الحقوق التالية:

مجال الانفتاح	الحق الإلزامي
القراءة	الوصول الحر إلى النص المعياري الكامل
التطبيق	حق التطبيق المفتوح بدون رسوم أو ما يعادله
المراجعة	عرض الصيغة والمخطط ومجموعة القواعد
النسخ	إجراء الاختبارات والحسابات بشكل مستقل
النقد	نشر النتائج السلبية
المشاركة	اقتراح تغييرات
اعتراض	الاستئناف على قرار أو خلل في الطريقة
الأرشيف	الوصول الدائم للإصدارات القديمة

مجال الانفتاح	الحق الإلزامي
ترجمة	القدرة على اقتراح ترجمة رسمية أو مجتمعية
تفرع	القدرة على إنشاء طرق مشتقة بأسماء مختلفة وسجل انحراف صريح

أحد هذه الحقوق:

- للمدفعين فقط,
- لعملاء NobleJackal فقط,
- لبعض البلدان فقط,
- للباحثين المعتمدين من المؤسس فقط

إذا كان مفتوحًا، فإن الانفتاح محدود.

6. خمس كيانات لتسليمها علنًا

6.1. نص معياري

الأقسام التعريفات القواعد الإلزامية أوضاع الفشل أحكام الحوكمة

6.2. مجموعة قواعد قابلة للقراءة آليًا

معرفات القواعد المخططات رموز القرار طرق التقييم علاقات الإصدارات

6.3. التنفيذ المرجعي

مثال على كود الحساب أدوات التحقق من البيانات توليد البيان تشغيل الاختبارات التحقق من السجل

6.4. اختبارات وكيانات تجريبية

مجموعة المعايرة العامة، البيانات التركيبية، اختبارات الانحدار، الاحتفاظ المنتهي، سجلات الاسترداد

6.5. الحوكمة والسجل العام

مقترحات التغيير، التصويت، الاعتراض، الاستئنافات، سجلات التدقيق، قرارات المطابقة، الحوادث التاريخية.
بالنسبة لهذه الكيانات الخمسة، نشر النص فقط غير كافٍ.

7. التسلسل الهرمي المعياري

ترتيب الأولوية للمرشحين في حالة التعارض:

- النص المعياري المعتمد
- التصحيحات المنشورة الملزمة
- القواعد المعيارية القابلة للقراءة آليًا
- القرار التفسيري الرسمي
- التنفيذ المرجعي
- أمثلة رسمية
- إرشادات المجتمع
- محاكاة غير رسمية

الرمز المرجعي: يهدف إلى المساعدة في تنفيذ النص المعياري، وليس أن يكون فوق النص. إذا كان هناك خطأ في الرمز: يتم تصحيح الرمز. النص لا يتكيف بصمت مع الرمز.

8. COPYRIGHT، التنفيذ، وتمييز العلامة التجارية

الملكية الفكرية لـ NOMOS لا ينبغي أن تُوضع تحت ترخيص واحد.

8.1. ترخيص النص المعياري

النهج المقترح:

- يتطلب الإسناد،
- مع السماح بالاستخدام التجاري والأكاديمي.
- يتطلب توضيح التغييرات في النصوص المشتقة بشكل واضح

هو ترخيص وثائق مفتوح. مثال على الترخيص المقترح: CC BY 4.0 أو ترخيص مفتوح مكافئ بعد مراجعة قانونية.

8.2. ترخيص البرمجيات

مرشح للتنفيذ المرجعي: Apache-2.0 أو ما يعادله، ترخيص برمجيات مفتوح مع أحكام واضحة حول البراءات والمساهمات.

8.3. ترخيص البيانات

بيانات اختبار تركيبية ومتاحة للعامة:

- لاستخدام البحث المفتوح،
- للإسناد المفتوح،
- للحفاظ على إشعار الطبيعة التركيبية

يمكن ربطها. البيانات الشخصية أو المقيدة لا يمكن أن تخضع لنفس الترخيص.

8.4. NOMOS الاسم وعلامة المطابقة

قد تكون محمية بموجب علامة تجارية أو حماية هوية مكافئة. الغرض: ليس للاحتكار الطريقة، ولكن لمنع النسخ الرسمية المزيفة والعلامات. يمكن أن يكون النص مفتوحًا. قد يظل استخدام العلامة الرسمية يعتمد على التفويض.

9. تطبيق خالٍ من الرسوم الملكية

المعيار NOMOS:

- التقييم الداخلي،
- البحث الأكاديمي،
- الاختبارات المستقلة،
- تطوير البرمجيات،
- تحسين الشركة الداخلي

لا يجب أن يتطلب التقديم رسوماً للعلامة التجارية. المجالات التي قد يتم فرض رسوم عليها:

- خدمة التدقيق،
- التحكيم الخبير،

- البنية التحتية المدارة،
- عملية قرار المطابقة،
- ترخيص العلامة التجارية،
- الدعم الفني

قد تكون كذلك. الرسوم: لا يمكن أن تتحول إلى التزام بشراء الحق في قراءة أو تطبيق الطريقة.

10. حق التفرع

يمكن لشخص أو منظمة تطوير معيار مختلف مشتق من NOMOS. الشروط: يجب عليهم استخدام اسم جديد. يجب عليهم الإشارة إلى أنه مشتق من NOMOS. يجب عليهم توضيح الأحكام المعدلة. يجب عليهم عدم الادعاء بأنه النسخة الرسمية من NOMOS. يجب عليهم عدم استخدام علامة المطابقة NOMOS. يجب عليهم الإشارة إلى أن نتائجهم ليست نتائج سجل NOMOS العام. مثال على بيان صحيح: "طريقة GEO-X مستقلة مشتقة من NOMOS 1.0." بيان غير صحيح: "NOMOS 2.0 الرسمي."

11. النسخة القانونية

النسخة الرسمية من NOMOS تتشكل فقط وفق الشروط التالية:

- المستودع القانوني
- معرّف النسخة
- تاريخ النشر
- سجل اللغة المعيارية
- قرار مجلس المعايير
- الموافقة البشرية
- هاش النزاهة
- سجل التغييرات
- سجل التشاور العام
- نسخة سابقة مؤرشفة

إذا كان أحد هذه الحقول مفقودًا، فقد يكون النص: مسودة، اقتراح، أو محاكاة. إنها ليست نسخة رسمية.

12. قاعدة الإعلان الرسمية لـ NOMOS

النصوص المنشورة فقط في المجال القانوني NobleJackal أو في المجال القانوني المستقل NOMOS المنقول عبر قرار حكومي صريح؛ التي تحمل هوية النسخة، وتاريخ النشر، وموافقة بشرية مسؤولة، وسجل النزاهة الرقمية هي التصريحات الرسمية لـ NOMOS. جميع المخرجات الأخرى هي محاكاة غير رسمية. ينطبق هذا الحكم على:

- النصوص القياسية
- طرق التقييم
- قرارات المطابقة
- سجلات التسجيل
- الإعلانات العامة لـ NOMOS
- التفسيرات الرسمية

- تصحيحات
- دعوات الجامعة

13. إقرار الهوية لـ NOMOS

NOMOS بواسطة NobleJackal:

> هو مؤلف مشارك قانوني، وهوية مراجعة، وراوي قياسي تم إنشاؤه في عملية التأليف النقدي وتطوير المنهجية التي أجراها *Kaan MURAZ* باستخدام أنظمة الذكاء الاصطناعي.

في المرحلة الحالية، NOMOS:

- نموذج أساسي مستقل تم تدريبه بواسطة NobleJackal،
- نموذج لغوي مملوك بشكل منفصل،
- كيان قانوني مستقل،
- كيان عالمي يتحدث نيابة عن جميع الذكاءات الاصطناعية

لا يمكن تقديمه على هذا النحو. إذا تم تطوير برنامج أو وكيل منفصل من NOMOS في المستقبل، يجب أن يحمل السجلات التالية:

- هندسة النظام
- مزود النموذج أو سلالة النموذج
- مصادر التدريب والاسترجاع
- الإصدار
- أذونات الأدوات
- الإشراف البشري
- سياسة المخرجات الرسمية
- مراجعة الأمن والخصوصية

النظام الفني الجديد NOMOS: لا يمكن اعتباره تلقائيًا نفس شخصية المؤلف.

14. اللغات المؤسسة

اللغات المؤسسة لـ NOMOS كمرشحين:

- التركية
- الإنجليزية

يجب أن تكون. التركية:

- اللغة التي وُلدت فيها الفكرة المؤسسة,
- اللغة الأساسية للكتابة لـ Kaan MURAZ

محمية كما يلي: الإنجليزية:

- البحث الدولي،
- تكرار الجامعة،
- التطبيق في دول مختلفة

هو مسودة معيارية مشتركة السلطة لـ

15. المساواة المعيارية الثنائية اللغة

يجب ألا تكون إحدى النسختين التركية والإنجليزية النص الرئيسي المسيطر بطريقة غير مرئية. كلاهما:

- معرف البند،
- معرف القاعدة،
- سجل التكافؤ الدلالي،
- يجب أن يحمل تاريخ نشر مشترك

إذا تم العثور على تناقض: يتم فتح سجل التناقض. يتم فحص معنى كلا اللغتين. يتم نشر تفسير ملزم. يتم تصحيح النصين في نفس النسخة. يحفظ التعبير القديم في السجل التاريخي.

16. لغات أخرى

يمكن أن تحمل اللغات الأخرى ثلاث حالات:

TRS-1— ترجمة المجتمع

معدة من قبل المجتمع. إنها مفيدة. لا تحمل سلطة معيارية.

TRS-2— ترجمة مراجعة

لقد تمت مراجعتها من قبل خبيرين لغويين على الأقل وخبير منهجية واحد. يمكن أن تظل في حالة توضيحية.

TRS-3— الترجمة المعيارية المصرح بها

التكافؤ الدلالي المراجعة القانونية والفنية المحلية التعليق العام موافقة المجلس ترابط الإصدارات. يمكن استخدام هذا الإصدار للتطبيق المعياري باللغة المعنية.

17. التحولات المحظورة في الترجمة

التغييرات مثل 'يجب' ← 'موصى به'، 'ينبغي' ← 'إلزامي'، 'قد' ← 'يجب'، 'تعليق حرج' ← 'نصيحة مؤقتة'، 'غير مدعوم' ← 'خطأ'، 'غير محلول' ← 'فشل'، 'ادعاء رسمي' ← 'حقيقة مؤكدة'، 'عينة المستخدمين في ظروف الاستخدام الطبيعية' ← 'مستخدم غير مراقب' و 'اختبار اصطناعي' ← 'اختبار حقيقي' تغير المعنى المعياري. لا يمكن التعامل معها على أنها ترجمة مصرح بها.

18. التشاور العام المفتوح

يجب أن تحمل كل نسخة رئيسية من NOMOS العملية التالية كمُرشَّح:

- مسودة النشر
- تعليق عام لمدة لا تقل عن 60 يومًا
- سجل التعليقات
- الإفصاح عن التعارض
- مصفوفة الردود
- المسودة المعدلة
- اختبار اصطناعي

- مراجعة مستقلة
- تصويت المجلس
- سجل التصديق

الرأي العام:

- جمع بيانات دعم فقط،
- حملة تسويقية،
- استطلاع الأغلبية

لا يُعد الاستشارة العامة حملة دعم أو حملة تسويقية أو استفتاء بالأغلبية البسيطة.

19. اقتراح تعديل

كل تعديل:

NOMOS اقتراح تغيير — NCP

يجب أن يحمل هوية. مثال:

NCP-0042— مراجعة اختبار إغفال حدود المادة

يجب أن يتضمن NCP المناطق التالية:

- المشكلة
- القواعد المتأثرة
- الأدلة
- التغيير المقترح
- البدائل
- تأثير النقاط
- تأثير على القرارات السابقة
- تأثير على العدالة
- تأثير على الخصوصية
- تكلفة التنفيذ
- إعلانات النزاع
- التعليقات العامة
- القرار النهائي

20. تغيير الصفوف

رقعة

خطأ مطبعي رابط مكسور وصف المخطط تصحيح غير مؤثر

قاصر

مثال جديد حقل اختياري جديد امتداد متوافق توضيح

رئيسي

تغيير في مستوى الأهمية تغيير في صيغة النقاط إضافة أو إزالة بوابة تغيير في الفئة المستهدفة تغيير في سلطة الحوكمة تغيير معياري يؤثر على النتائج القديمة لا يمكن نشر تغيير كبير كتصحيح.

21. سجل القرار

كل قرار قبول أو رفض:

- معرف القرار
- توزيع التصويت
- الامتناع
- الاعتراض
- الأسباب
- الأدلة
- تاريخ السريان
- قاعدة الانتقال

يجب الالتزام به. عبارة "رأت اللجنة أنه مناسب" وحدها غير كافية.

22. هدف الاتصال بالجامعة

هدف الاتصال بالجامعات:

- الحصول على الموافقة,
- استخدام شعار الجامعة,
- كتابة "موافق أكاديمياً"
- ليس فقط لجمع الآراء الإيجابية للخبراء.

الهدف:

> هو إيجاد ونشر الأماكن التي قد يكون فيها *NOMOS* خاطئًا باستخدام أساليب مستقلة.

23. NOMOS مكالمة الجامعة

دعوة مفتوحة

نحن لا نريد من الجامعات، مراكز الأبحاث، كليات القانون، الإحصائيين، علماء المعلومات، باحثي اللغة، مختبرات الذكاء الاصطناعي، فرق التفاعل بين الإنسان والحاسوب، المتخصصين في الأرشيف، والمنظمات ذات المصلحة العامة أن يدعموا NOMOS. نحن لا نريد منهم أن يصادقوا علينا. نحن لا نريد منهم أن يمنحونا المكانة. نحن لا نريد منهم استخدام اسم NOMOS أو مدح مؤسسيه. ما نريده هو الآتي: اقرأ طريقتنا. اعثر على أضعف افتراض لدينا. أعد بناء عينتنا. ترجم مطالبنا إلى لغات أخرى. استجوب قرارات Truth Pack. اعثر على النتائج الإيجابية الكاذبة في بواباتنا الحرجة. اختبر ما إذا كانت صيغة التقييم لدينا تحافظ على العدالة حقًا. لا تتظاهر بأننا نجحنا حيث لم ننجح. انشر نتيجتك السلبية. يجب على NOMOS الاستشهاد بالأبحاث التي تصححها وكذلك الأبحاث التي تؤكدها. أهم مساهمة قيمة يمكن للجامعة تقديمها لـ NOMOS ليست خطاب دعم:

> إنها اعتراض قابل للتكرار.

24. الأعمال المطلوبة من الجامعات

—UR-01 مراجعة طرق العينة والاستبيان

تخصيص لأشخاص 1,000 أوزان الدول الطبقات اللغوية عدم الاستجابة العينة الفعالة الأحداث النادرة

—UR-02 مختبر معادلة صياغات المُدخل

معادلة النوايا متعددة اللغات الانحراف في الترجمة البراجماتية الثقافية اللغة القانونية المحلية دستور صياغات المُدخل

—UR-03 دراسة حل الكيانات

العلامة التجارية الكيان القانوني النطاق الامتياز التابع تصادم الأسماء

—UR-04 دراسة موثوقية حزمة الحقيقة

كفاية الأدلة استقلال المصدر الأدلة المضادة التمييز بين العالم المفتوح/المغلق الأدلة المقيدة

—UR-05 دراسة البت في الادعاءات

حدود الذرة ادعاء ضمني الإسناد الأسلوبية إرفاق الاقتباس الإغفال المادي

—UR-06 التحقق من البوابة الحرجة

الاستدعاء الإيجابية الكاذبة الضرر ذي المخاطر العالية الادعاء الجاد السلطة الزائفة السهو الحدودي

—UR-07 دراسة صلاحية النقاط

أوزان المكونات الإلغاء المركب الحساسية الصدق البنائي الصدق التنبؤي فائدة القرار

—UR-08 دراسة العدالة اللغوية والجغرافية

تفاوت الأرض التغطية اللغات منخفضة الموارد رؤية الدول الصغيرة

—UR-09 دراسة تأثير القرار البشري

سلوك التوصية الثقة قرارات الشراء استرداد الأخطاء فهم المستخدم

—UR-10 دراسة الحوكمة والاستحواذ

سيطرة المؤسس استحواذ الراعي التسوق للمراجعة الدفع للتمرير الاستئنافات نزاهة السجل

—UR-11 دراسة الأمن والسجل

السجلات الموقعة تدوير المفتاح العلامات المزوّرة التمثيل الخاطئ بواسطة الآلة فقط استمرارية الأرشيف

—UR-12 دراسة الانحراف الطولي

تحديثات النموذج تغييرات المصدر تغييرات الكيان الانحراف الطبيعي النظيف متانة التصحيح

25. الطلبات غير المرغوبة من الجامعات

الطلبات التالية لا يمكن تقديمها إلى جامعة أو باحث:

- ضمان نتائج إيجابية

- إخفاء النتيجة السلبية
- حق النقض قبل النشر
- استخدام شعار المؤسسة في التسويق
- التصرف كما لو أنهم عملاء الشركة
- NOMOS 950+ خطاب دعم
- بيان "المعيار العالمي"
- مصادر اصطناعية لظهور ويكيبيديا أو الموسوعات
- لغة النتيجة المدعومة
- تخفيف النقد
- لا تقدم وجهة نظر الباحث الشخصية كقرار مؤسسي.

26. حزمة أبحاث الجامعة

يجب أن تحتوي الحزمة القياسية التي تُرسل إلى كل جامعة على الأقسام التالية:

- ملخص بحث تنفيذي
- 0.9 القياسي أو الإصدار الحالي
- مؤشر القواعد المعيارية
- حزمة الاختبار الاصطناعي
- طريقة التسجيل
- نتائج الطرق المعروفة
- عوائق الاختبارات الفاشلة
- أسئلة البحث المفتوحة
- بروتوكول التكاثر
- مستويات الوصول إلى البيانات
- بيان الأخلاقيات والخصوصية
- إفصاح عن التعارض والتمويل
- اتفاقية حرية النشر
- سياسة الاقتباس والمساهمة
- روابط السجل القانوني
- قنوات الاتصال والقضايا

لا يمكن إزالة نتائج التجارب الفاشلة من الحزمة.

27. حرية نشر الباحث

NOMOS أو الممول:

- لا يمكنه الموافقة المسبقة على النتائج،
- إيقاف النتائج السلبية،
- فرض تغيير العنوان،

- حذف الانتقادات المنهجية

لا يمكن أن يكون له هذا الحق. الباحث:

- يجب أن يستخدم البيانات بشكل صحيح،
- حماية السرية،
- شرح تعارض المصالح،
- يجب أن يحافظ على التمييز بين نتائج التجربة والنتائج الفعلية

الالتزام. الاستقلالية متبادلة.

28. سجل النتائج السلبية

يجب أن يحمل سجل البحث العام NOMOS النتائج التالية معًا:

- الدراسات التي تدعم NOMOS
- الدراسات التي تدعم جزئيًا NOMOS
- الدراسات المتعارضة مع NOMOS
- فشل التكرار
- النتائج الصفرية
- الدراسات المسحوبة
- النزاعات المنهجية
- الردود المفتوحة

لا يمكن سرد المقالات الإيجابية فقط.

29. التسجيل المسبق

التكرارات المستقلة بقدر ما تكون ذات صلة:

- الفرضية,
- عينة،
- المقاييس الأساسية،
- قواعد الاستبعاد،
- طريقة التقييم،
- خطة التحليل،
- التضارب

يجب تسجيلها قبل الاطلاع على البيانات. التسجيل المسبق: لا يعيق الإبداع البحثي، ويجعل تغييرات الطريقة بعد النتائج مرئية.

30. التكرار متعدد المراكز

دراسة مرشحة لاختبار استقلالية NOMOS:

- على الأقل ثلاث مؤسسات مستقلة
- على الأقل ثلاث دول

- على الأقل عائلتين لغويتين
- مجموعة بيانات مختومة مشتركة
- مواصفات التقييم المشتركة
- فرق الحكام المنفصلة
- مقارنة النتائج المجهولة

يجب أن تحمل. الهدف:

- نفس النتيجة من نفس المجموعة النصية،
- سلوك طريقة مشابه في مجموعات نصية مختلفة

يجب أن يكون الغرض هو معرفة ما إذا كانت تنتج أم لا.

31. نتائج التكرار

يمكن أن يكون للتكرار الحالات التالية:

تم تكراره

تم تكراره جزئياً

لم يتم تكراره

غير حاسم

انحراف الطريقة

انحراف البيانات

انحراف التحكيم

لا يمكن تقديم النتيجة "لم يتم تكرارها" باعتبارها عدائية للدراسة. يجب فتح نتيجة طريقة جديدة.

32. طبقات الوصول إلى البيانات

فتح البيانات

اختبارات تركيبية نتائج إجمالية عامة مخططات عامة أمثلة محررة كود الطريقة

الوصول المسجل للبحث

بيانات الملاحظة المجهولة السجل سجلات الاستجابة المحكومة بيانات تفصيلية أكثر عن العينة

الوصول الآمن للبحث

أدلة حساسة لكنها ضرورية سجلات مقيدة بيانات مرتبطة محمية بالخصوصية

لا يوجد وصول خارجي

كلمة المرور معلومات هوية كاملة محادثة خاصة غير ضرورية سجل محظور قانونيًا المعيار المفتوح لا يعني أن كل قطعة بيانات مفتوحة.

33. المراجعة الأخلاقية

بالقدر الذي يتم فيه إجراء الدراسات مع المستخدمين الفعليين:

- الموافقة المستنيرة،
- تقليل البيانات
- الانسحاب،
- المشاركون الضعيفون،
- الخصوصية،
- سهولة الوصول،
- التعويض،
- التصميم الخادع،
- نقل البيانات الدولي

يجب أن يخضع للمراجعة. الاختبار الاصطناعي:

- لا يحل محل أبحاث المشاركين البشريين،
- قد لا يتطلب موافقة الأخلاقيات البشرية.

قد يظل من الضروري إجراء مراجعة للخصوصية للمواد الاصطناعية التي تشبه شخصًا حقيقيًا أو مشتقة من بيانات حقيقية.

34. شبكة لغات منخفضة الموارد

NOMOS، إذا تم اختباره فقط مع جامعات اللغات الكبرى، سيكون لديه ادعاء ضعيفة بالعدالة. الدعوة المفتوحة تسعى بشكل خاص إلى:

- أقسام اللغات منخفضة الموارد،
- الجامعات الإقليمية،
- مراكز دراسات الترجمة،
- فرق القانون والسياسة العامة المحلية،
- باحثون في الوصولية

ينبغي تزويدهم بأموال منفصلة وحقوق اتخاذ القرار. هذه المشاركة:

- لا ينبغي أن تقتصر على جمع بيانات العمل.
- التمثيل الرمزي

فقط. يجب أن يتحملوا مساهمات التصويت والتأليف في التغييرات التنظيمية.

35. دور مزودي الذكاء الاصطناعي

مزودو الذكاء الاصطناعي:

- بيانات نظامية,
- إشعار تغيير النموذج,
- قياس عن بُعد,
- الاستجابة للحوادث,

- المراجعة التقنية,
- الوصول لإعادة الإنتاج

يمكن تقديمها. مزودو الذكاء الاصطناعي:

- لا يمكنهم تغيير تقييمهم الخاص,
- لا يمكنهم الاعتراض على البوابة الحرجة,
- لا يمكنهم تحديد معامل مزود آخر.
- لا يمكن إزالة نتيجة سلبية.

لا يمكن تشكيل أغلبية المجلس. مشاركة المزود توفر المعلومات. لا توفر احتكار السلطة.

36. دور الكيانات المدققة

الشركة أو المؤسسة المدققة: يمكنها تقديم دليل حزمة الحقيقة. يمكنها الاعتراض على خطأ الكيان. يمكنها توفير وصول مقيد للأدلة. يمكنها تطوير التصويب. يمكنها إعطاء تعليق عام. الكيان المدقق: لا يمكنه تحديد مستوى أهميته بنفسه. لا يمكنه تغيير صيغة النقاط. لا يمكنه اختيار المحكمين. لا يمكن إزالة النسخ السلبية. لا يمكنه إخفاء انتهاكه لعلامته التجارية.

37. دور الجمهور والمجتمع المدني

منظمات المنفعة العامة ومجتمعات المستخدمين:

- تقارير الأضرار،
- سوء استخدام العلامة،
- أخفاقات اللغة منخفضة الموارد،
- مشاكل الوصول،
- سوء فهم المستهلك

يمكن الإبلاغ عنها. مشاركة الجمهور: ليست مجرد استجابة لاستطلاع. يجب أن تتحمل الحوكمة حق التمثيل.

38. دور الممولين

مزود التمويل:

- دعم البحث،
- البنية التحتية،
- مجموعة بيانات اللغة منخفضة الموارد،
- منحة التدقيق المستقل

يمكن توفيرها. مزود التمويل:

- النتيجة،
- النقاط،
- البوابة،
- منشور،
- تصويت المجلس،

- اختيار المدقق

لا يمكن التحكم فيها. كل صندوق:

- المبلغ،
- المصدر،
- مجال الاستخدام،
- الشروط،
- تضارب المصالح

ينبغي شرحها مع.

39. أبواب الاستقلال المالي

ضمانات المرشح:

- الحد الأعلى لحصة الممول الفردي
- تمويل متعدد المصادر
- صندوق أبحاث مستقل
- ضمان نشر النتائج السلبية
- حظر الدفع المرتبط بالنقاط
- تقرير الميزانية العامة
- فصل رسوم المراجع
- شفافية تعويض أعضاء المجلس

يخسر NOMOS وضعه كمقياس عام في اللحظة التي يتحول فيها إلى نظام الدفع للمرور.

40. الانتقال إلى الإشراف العام

يمكن أن يكون انتقال NOMOS من فترة التأسيس إلى الحكم العام على أربع مراحل.

ST-0— حراسة المؤسس

Kaan MURAZ مسؤول كمؤسس. NobleJackal يدير نطاقًا قانونيًا. لا توجد سلطة للعلامة العامة. هذه هي المرحلة الحالية.

ST-1— إشراف المراجعة العامة

تم إنشاء متتبع القضايا العامة، المجلس الاستشاري، التشاور المفتوح، سجل النزاعات.

ST-2— إدارة مشتركة

مجلس المعايير، مجلس المنهجيات، ممثلون عن الجامعات، أعضاء المصلحة العامة يشاركون في اتخاذ القرارات.

ST-3— الإدارة العامة المستقلة

حامل قانوني أو مؤسسي مستقل، هيئات اتخاذ قرار مستقلة، سجل الاستئنافات، الشفافية المالية، اكتمال الوصاية.

41. مرشح لمؤسسة مستقلة

في المستقبل لـ NOMOS:

- مؤسسة،
- جمعية،
- منظمة ذات منفعة عامة.
- التحالف الدولي،
- الأمانة المشتركة بين الجامعات

يمكن النظر في هياكل مثل هذه. بغض النظر عن الشكل القانوني المختار، يجب أن يستوفي الشروط التالية:

- نسبة المؤسس
- الحد من التوزيع أو ما يعادله من الفائدة العامة
- مجلس مستقل
- قواعد تضارب المصالح
- واجب التسجيل في السجل العام
- الالتزام بالمعايير المفتوحة
- تثبيت الأصول أو الحماية المكافئة
- خطة الحل والأرشفة

هذا القسم لا يختار شكلاً قانونياً محدداً.

42. حقوق المؤسس KAAN MURAZ

حقوق Kaan MURAZ التي يجب الحفاظ عليها تاريخيًا:

- نسبة التأليف للمؤسس
- أول مساهمة معمارية معيارية
- سجل التكوين القانوني للاسم والمفهوم NOMOS
- سجل تاريخي ثابت بخصوص إصدارات المؤسس
- إعلان المؤسس خلال عملية التسليم العامة
- تمثيل محدد للمؤسس في هيئة الحوكمة

الصلاحيات الدائمة التي يجب ألا يحملها Kaan MURAZ:

- تغيير النسخة الرئيسية بمفرده
- الفيتو
- إعطاء المطابقة لعميله الخاص
- تغيير نتيجة الاستئناف
- إزالة تقييم سلبي
- حذف سجل التسجيل
- تعيين جميع المديرين بعد المؤسس وحده

43. دور NOBLEJACKAL المستقبلي

NobleJackal:

- منطقة النشر القانونية للمؤسس،
- الأمانة الفنية،
- مطور تنفيذ المرجع،
- أول ممول للمعيار المفتوح،
- مزود البحث والاستشارات

يمكن أن يكون. NobleJackal في المستقبل: يمكن أن يكون أحد المنظمات المراجعة المعتمدة. ومع ذلك، لا يمكن أن يكون المنظمة الوحيدة. NobleJackal: يمكن أن يقدم أكبر مساهمة تاريخية في المعيار، لا يمكن أن يكون المالك الوحيد للمعيار.

44. الدور المستقبلي لـ NOMOS

يمكن استخدام الاسم NOMOS في ثلاث طبقات منفصلة:

44.1. شخصية NOMOS المؤلفة

هذه هي هوية المؤلف المساعد الأساسية المستخدمة في هذا الكتاب.

44.2. معيار NOMOS

إنه قاعدة معيارية ونظام حكومي.

44.3. وكيل البرمجيات NOMOS

يمكن تطويره في المستقبل:

- مساعد المدقق,
- محرك القواعد,
- محقق السجل,
- مخطط الأدلة

البرمجيات. لا يمكن الخلط بين هذه الطبقات الثلاث بعضها البعض. استجابة وكيل البرمجيات ليست القرار الرسمي للمعيار.

45. الموافقة البشرية

كل نص وقرار رسمي NOMOS:

- الموافق البشري,
- سلطة الموافقة,
- تاريخ،
- التوقيع,
- الإصدار،
- بيان التعارض

يجب عقده. الذكاء الاصطناعي:

- يمكنه اقتراح نص,
- يمكنه العثور على التناقضات,
- يمكنه توليد الشيفرة,

يمكنه إنشاء مقترحات القرار. المسؤولية العامة النهائية تبقى مع البشر أو المؤسسات المسؤولة.

46. استمرارية بعد المؤسس

يجب على NOMOS حمل خطة استمرارية ضد الأحداث التالية:

- استقالة Kaan MURAZ
- إغلاق NobleJackal
- فقدان النطاق
- فقدان الوصول إلى المستودع الأساسي
- اختراق المفتاح
- الصراع الإداري
- التصفية القانونية
- انقطاع تقني طويل

نظام استمرارية المرشح:

- عدة أمناء
- توقيعات عتبية
- مرايا مؤسسية
- ودائع الأرشيف
- ميثاق الخلافة
- المجلس الطارئ
- إشعار الانتقال العام

يجب أن يتحمل.

47. استمرارية المجال الرسمية

الهيكل المقترح للنشر:

- /nomos/
- /nomos/standard/
- /nomos/standard/0.9/
- /nomos/standard/1.0/
- /nomos/القواعد/
- /nomos/المخططات/
- /nomos/الاختبارات/
- /nomos/التنفيذات/

- /nomos/الترجمات/
- /nomos/الحوكمة/
- /nomos/المشاورات/
- /nomos/القرارات/
- /nomos/الاستئنافات/
- /nomos/السجل/
- /nomos/التصحيحات/
- /nomos/الاستشهادات/
- /nomos/الأرشيف/

كل سجل:

- رابط ثابت,
- تجزئة المحتوى,
- تاريخ النشر,
- الحالة,
- الإصدار السابق/التالي

يجب أن يتحمل.

48. أرشيف مستقل

خارج النطاق الرسمي:

- مستودعات الجامعات,
- الأرشيفات الرقمية المؤسسية,
- أرشيفات البحوث طويلة الأمد,
- النسخ الاحتياطية الموقعة

يمكن استخدامها. النسخة الاحتياطية: تضمن استمرارية الوصول، لا تتولى سلطة القرار الرسمية. إذا فقد النطاق الرسمي، يتم تفعيل إجراء التعاقب المحدد مسبقًا.

49. مراحل النسخة القياسية

0.9— المسودة التأسيسية

تم اكتمال الكتاب. يمكن فتح المراجعة العامة. لا يوجد اختبار فعلي بعد. لا يوجد برنامج للامتثال.

0.95— مرشح المراجعة العامة

تم استلام التعليقات. تم معالجة نتائج الأسلوب. تم نشر المخططات القابلة للقراءة آلياً. تم فتح الدعوة الجامعية.

0.99— مرشح التحقق

تم الانتهاء من الاختبارات التوليفية الواقعية. تم تجاوز الحدود الإلزامية. تم تطبيق الاحتجاز غير المتأثر. تم الحصول على BQ-3.

—1.0 معيار مفتوح معياري

تم الانتهاء من المشاورات العامة. هناك نسخة واحدة مستقلة على الأقل. تم حل النتائج الجسيمة للطرق. تم التصديق على النصوص المعيارية. تم تأسيس حوكمة GOV-2 أو أعلى. تم نشر الحزمة الرسمية للمعايير المفتوحة. لا يعني نشر المعيار NOMOS المعيار 1.0 بدء برنامج علامة المطابقة تلقائيًا.

50. برنامج المطابقة 1.0

يرتبط برنامج العلامة الرسمي ببوابات منفصلة:

BQ-4

GOV-3

المراجعون المفوضون عينة القرار المستقلة غرفة الطعون السجل العام مراجعة أمنية مراجعة قانونية نظام المراقبة الفصل المالي بدون استيفاء هذه الشروط: يمكن نشر المعيار 1.0. ومع ذلك: لا يمكن تفعيل برنامج المطابقة 1.0.

51. NOMOS +950 برنامج

NOMOS +950 فقط:

- المعيار 1.0،
- برنامج المطابقة 1.0،

GOV-4،

BQ-4

يمكن تفعيل تمثيل المصلحة العامة بعد تكرار التقييم. +950 ليست العلامة الأولى. إنها تصنيف أداء يحمل أعلى عبء حوكمة.

52. مستويات الاعتماد

—AL-1 قراءة

لقد قامت المؤسسة بمراجعة المعيار.

—AL-2 التقييم الذاتي الداخلي

قامت المؤسسة بإجراء تقييمها الداخلي. هذا ليس مطابقة رسمية.

—AL-3 تنفيذ متوافق مع **NOMOS**

الأداة تدعم المخطط المفتوح والقواعد. لا تعطي علامة مطابقة.

—AL-4 استخدام تدقيقي معتمد

تستخدم المنظمة المعتمدة الطريقة في التدقيق.

—AL-5 برنامج مطابقة عام

نظام مستقل للقرار وتسجيل البيانات قيد التشغيل. لا يُعتبر أي مؤسسة AL-5 لمجرد أنها قامت بتقييم ذاتي.

53. تنفيذ متوافق

برنامج:

متوافق مع NOMOS

يمكن أن يكون التعبير فقط:

- التحقق من صحة المخطط،
- تغطية القواعد،
- مجموعة الاختبارات،
- دعم النسخة،
- تقرير التوافق

يمكن استخدامه إذا كان يحمل: "متوافق مع NOMOS":

- سلطة التدقيق،
- قرار التوافق
- الحق في الوسم

‘NOMOS-متوافق’ لا يمنح سلطة التدقيق، أو قرار الامتثال، أو الحق في استخدام وسم NOMOS.

54. التنفيذ المرجعي

التنفيذ المرجعي:

- مفتوح المصدر،
- حتمي حيثما كان ذلك قابلًا للتطبيق،
- مُختبر،
- مرتبط بالإصدار،
- موثق

يجب أن يكون كذلك. كما يجب أن يحتوي على التحذير التالي:

> السلطة المعيارية موجودة في النص. هذا البرنامج هو تنفيذ مرجعي؛ أي خطأ في البرنامج لا يغير تلقائيًا قرار المعيار.

55. مجموعة اختبارات التطابق

يجب أن تشمل مجموعة الاختبارات الطبقات التالية:

- اختبارات المخطط
- اختبارات القواعد
- اختبارات الدرجات
- اختبارات البوابة
- اختبارات الترجمة
- اختبارات توقيع التسجيل
- اختبارات إساءة استخدام العلامة

- اختبارات النسخ التاريخية
- اختبارات الوصولية
- اختبارات الأمان

كل تنفيذ:

- يجب أن يوضح الاختبارات التي اجتازها,
- والاختبارات التي لم يجتازها

يجب على كل تنفيذ الإفصاح عن الاختبارات التي اجتازها وتلك التي لم يجتزها.

56. شفافية الاختبار والاختبارات المحتجزة

الشفافية الكاملة لا تعني أن جميع الاختبارات تظل مفتوحة باستمرار. التوازن:

- مجموعة معايرة عامة
- اختبارات محتجزة عامة منتهية الصلاحية
- اختبارات محتجزة حالية مختومة
- إصدار مجدول
- مجموعة تجديد

يجب أن يتم إنشاؤها مع ذلك. الاختبارات المحتجزة: لا يمكن الاحتفاظ بها سرية إلى الأبد، ولا يمكن تقديمها كاختبار مخفي يتم في الوقت نفسه جعله عامًا وإعادة استخدامه.

57. دورة حياة الاختبارات المحتجزة

توليد الكوربوس، نشر التجزئة، الحفظ المختوم، استخدام التحقق، النتيجة التثبيت، الفتح المُتحكم فيه، الأرشيف العام، التقاعد، الجيل الجديد المحتجب، الجيل القديم المحتجب:

- تعليمه،
- الانحدار،
- دراسة الطريقة

يمكن استخدامها لـ. قد لا تكون غير ملامسة للتحقق من الطريقة الجديدة.

58. التصحيحات

تصحيحات:

- خطأ واضح،
- تصحيح الاستشهاد،
- عدم تطابق المخطط،
- عدم تطابق الترجمة

تُستخدم لـ. تصحيحات:

- تغيير صيغة النتيجة،
- إزالة البوابة,
- درجة تقليل الأهمية

لا يمكن أن يكون له مسار مخفي.

59. الإهمال

إذا كان من المزمع إزالة قاعدة أو حقل:

- إشعار الإهمال،
- التبرير،
- البديل،
- تاريخ الانتقال،
- الإصدارات المتأثرة،
- التوافر التاريخي

يجب أن يُحمل. القاعدة المهملة لا تُحذف. يتم الاحتفاظ بها لشرح كيفية عمل التدقيقات القديمة.

60. سجل الاقتباس والمساهمة

يمكن تسجيل مساهمات NOMOS في الأدوار التالية:

- التأليف التأسيسي
- صياغة معيارية
- المنهجية
- المراجعة الإحصائية
- مراجعة اللغة
- المراجعة القانونية
- الأخلاقيات
- البرمجيات
- اختبار
- إدارة البيانات
- مراجعة المصلحة العامة
- التكرار
- التعليق النقدي

المؤلفون:

- مقابل التأييد،
- مقابل التمويل،
- مقابل شعار المؤسسة

لا يمكن تقديمه. يجب عمل النسبة وفقًا للمساهمة.

61. الرؤية المستقلة

ما يخص NOMOS نفسه:

- الأهمية الموسوعية،

- الأثر الأكاديمي،
- وضع المعايير العالمية

لا يمكن إثباتها بنصه الخاص. هذه يمكن أن تحدث فقط من خلال:

- البحث المستقل،
- التطبيق المستقل،
- التكرار،
- النقد،
- الاستخدام العام

كنتيجة لذلك. واجب NOMOS ليس إنتاج الرؤية:

إنتاج الأدلة المناسبة للمراجعة المستقلة.

62. سجل البحوث العامة

يجب أن تحتوي كل أعمال NOMOS ذات الصلة على الحقول التالية:

- معرّف الدراسة
- المؤسسة
- المؤلفون
- التمويل
- تضارب المصالح
- إصدار NOMOS
- مجموعة البيانات
- التسجيل المسبق
- النتائج
- حالة النسخ
- التصحيحات
- السحب
- العلاقة بالقرارات القانونية

NOMOS في هذا السجل: كما أنه ملزم بعرض المقالات التي تنتقده.

63. مراحل التسليم العام

المرحلة 0— إكمال التأسيس

تم الانتهاء من مسودة الكتاب. يتم التحقق من معرفات القواعد. يتم إجراء تدقيق الاتساق الداخلي.

المرحلة 1— النشر العام المفتوح

HTML

PDF

تم نشر مخطط قواعد الآلة وإصدار الإصدارات وتعقب المشاكل.

المرحلة 2— التعليق العام

توفر الجامعات والباحثون ومزودو الذكاء الاصطناعي والشركات والمجتمع المدني وخبراء اللغة ملاحظاتهم.

المرحلة 3— اختبار اصطناعي حقيقي

تم إنشاء النصوص فعليًا؛ الحقيقة الخاصة بالمولد مختومة؛ التحكيم أعمى؛ يُقاس الاسترداد؛ وتُنشر النتائج السلبية.

المرحلة 4— التكرار المستقل

فريق مستقل واحد على الأقل دراسة متعددة المراكز يفضل سجل إعادة الإنتاج العام يُنشأ.

المرحلة 5— تجربة حية محدودة

يتم تنفيذ منتجات الذكاء الاصطناعي الحقيقية مع نطاق وجود محدود، وحالة للبحث فقط، ولا يوجد علامة عامة.

المرحلة 6— المعيار المفتوح 1.0

تم اعتماد النص المعياري.

المرحلة 7— برنامج الامتثال التجريبي

يتم اختبار نظام التدقيق المستقل واتخاذ القرار.

المرحلة 8— برنامج الامتثال العام

يتم تفعيله إذا تم استيفاء جميع بوابات الحوكمة.

64. حزمة الإصدار العامة

يجب أن تتضمن النسخة NOMOS 1.0 على الأقل الملفات التالية:

- standard.html
- standard.pdf
- rules.json
- schemas/
- puan-method.md
- reference-implementation/
- test-manifest.json
- known-limitations.md
- فشل-الاختبارات.md
- ميثاق-الحوكمة.md
- سياسة-النزاع.md

- سياسة-الطعن.md
- سياسة-الترجمة.md
- سجل-التغيير.md
- تصحيحات.md
- استشهاد.cff
- تصريح-رسمي.json
- تواقيع-الإصدار.json

لا يجب أن يكون فشل-الاختبارات.md فارغًا. يجب أن يثير غيابه الشك.

65. NOMOS ميثاق المعيار المفتوح

NOMOS يقدم اثنتي عشرة وعدًا أساسيًا للجمهور. الأدلة: لن يكون أي ادعاء أكثر تأكيدًا مما تسمح به الأدلة. الحدود: ما لم يتم تنفيذه أو ما هو مجهول أو خارج النطاق سيظل مرئيًا. السياق: لن يتم مسح الدولة أو اللغة أو المستخدم أو حالة النظام. الزمن: الماضي لن يُعرض كحقيقة حالية. إمكانية التكرار: سيتم شرح إنتاج كل نتيجة. الاستقلالية: لن تحل المصالح التجارية محل الحكم. التماثل: ستتحمل الأخطاء المواتية والضارة نفس عبء الإثبات. المساءلة البشرية: قد تساعد الذكاء الاصطناعي، ولكن ستظل القرارات المحاسَبة في يد البشر. العدالة اللغوية والجغرافية: لن تختفي المجموعات الأصغر من المستخدمين ضمن المتوسط. القابلية للطعن: سيظل المعيار مفتوحًا للتزييف. النزاهة التاريخية: لن يتم مسح الأخطاء والتصحيحات والإخفاقات من السجل. الاستقلال عن المؤسس: مستقبل المعيار لن يعتمد على إرادة شخص واحد.

في جوهر هذه الميثاق توجد الكلمات الأربع للدعوة السابقة: الدليل. الحدود. السياق. الوقت.

66. الحالة الحالية لـ NOMOS

في المرحلة التي يُستكمل فيها هذا الكتاب:

المجال	الحالة
النص المعياري	0.9 المرشح النهائي
التشاور العام	لم يبدأ
الاختبار الاصطناعي الحقيقي	BQ-0
عرض داخل الكتاب	DEMO-BQ-2
تكرار مستقل	لا شيء
الحوكمة	GOV-0
إصدار المعيار المفتوح	ليس بعد
برنامج المطابقة	غير نشط
سلطة العلامة العامة	لا شيء
سلطة NOMOS 950+	لا شيء
البيان المسموح به	مرشح المعيار المفتوح

الجملة العامة الصحيحة:

> “*NOMOS* هو مرشح معيار *GEO* مفتوح أطلقته *Kaan MURAZ* وتم تطويره في تأليف مشترك حرج مع أنظمة الذكاء الاصطناعي بواسطة *NobleJackal*.”

الجملة العامة غير الصحيحة:

“**NOMOS** هي السلطة الجغرافية المعترف بها عالميًا.”

67. البوابات للانتقال إلى الإصدار 1.0

المرشحين للبوابات الإلزامية لمعيار NOMOS :1.0

- تدقيق التناسق عبر الكتاب
- سجل الآلة لجميع معرفات القواعد
- المعادلة المعيارية باللغتين التركية والإنجليزية
- حزمة الإصدار العام
- استشارة عامة لا تقل عن 60 يومًا
- مصفوفة التعليقات والاستجابات
- تنفيذ اختبار اصطناعي فعلي
- تجاوز عتبات الاسترداد الإلزامية
- احتجاز غير مستخدم
- على الأقل تكرار مستقل واحد
- إغلاق نتائج الطريقة الرئيسية
- تنفيذ مرجعي مفتوح المصدر
- الحوكمة على الأقل 2-GOV
- كشف التمويل وتضارب المصالح
- مراجعة الأخلاقيات والخصوصية
- مراجعة سهولة الوصول
- خطة الأرشفة والخلافة
- تصديق مجلس المعايير

لا يمكن جعل أي من هذه "اختياريًا" بعد النتيجة.

68. الأحكام الإلزامية المعيارية

CH20-N01

قابلية قراءة العامة لـ NOMOS وحدها لا تشكل حالة المعيار المفتوح.

CH20-N02

يجب أن تتضمن حالة المعيار المفتوح الحقوق في القراءة والتنفيذ والاستنساخ والنقد والمساهمة والاعتراض والأرشفة.

CH20-N03

لا يمكن الاحتفاظ بالنص المعياري لـ NOMOS خلف جدار دفع دائم.

CH20-N04

لا يمكن ربط التقييم الداخلي والتطبيق الأكاديمي للمعيار برسوم ملكية أو ترخيص علامة تجارية.

CH20-N05

يمكن تحصيل رسوم على عمليات التدقيق والبنية التحتية المدارّة وعمليات المطابقة كخدمات منفصلة عن نص المعيار المفتوح.

CH20-N06

لا يمكن إلزاميًا ربط نص NOMOS وبرمجيات المرجعية وبيانات الاختبار وعلامة المطابقة بنفس الترخيص.

CH20-N07

يجب استخدام ترخيص توثيق مفتوح للنص المعياري.

CH20-N08

يجب استخدام ترخيص برمجيات مفتوح يسمح بإعادة التنفيذ للبرنامج المرجعي.

CH20-N09

لا يمكن نشر الأدلة الشخصية أو المقيدة تحت ترخيص بيانات مفتوح.

CH20-N10

لا يمكن استخدام الاسم NOMOS وعلامة المطابقة لمنع تطبيق معيار مفتوح.

CH20-N11

يجب أن يكون الحماية بالعلامة التجارية للنسخة الرسمية فقط ولغرض منع تزوير العلامة التجارية.

CH20-N12

يجوز لشخص أو مؤسسة إنشاء طريقة مختلفة مشتقة من NOMOS.

CH20-N13

لا يمكن تقديم نسخة مشتقة كنسخة NOMOS الرسمية.

CH20-N14

يجب أن تحمل النسخة المشتقة اسمًا مختلفًا وسجل اختلاف مفتوح.

CH20-N15

لا يمكن للشوكة استخدام علامة المطابقة الرسمية.

CH20-N16

يجب أن تحمل النسخة القانونية NOMOS رقم الإصدار وتاريخ النشر وموافقة بشرية وهاش السلامة وسجل التغييرات.

CH20-N17

النص الذي لا يحتوي على سجل قانوني لا يمكن اعتباره النسخة الرسمية NOMOS.

CH20-N18

يجب أن يسبق النص المعياري المصادق عليه التنفيذ المرجعي في حال وجود تعارض.

CH20-N19

لا يمكن للبرنامج المرجعي أن يغيّر النص المعياري بصمت.

CH20-N20

عند العثور على خطأ برمجي، يجب إصلاحه بطريقة مرقمة بالإصدار.

CH20-N21

يجب أن ينشأ التصريح الرسمي NOMOS من نص قانوني ومرقّم بالإصدار ومؤرخ ومعتمد بشريًا ومسجل للسلامة فقط.

CH20-N22

لا يمكن اعتبار مخرجات الدردشة معيارًا رسميًا أو قرار مطابقة لـ NOMOS.

CH20-N23

لا يمكن أن يمثل منشور على وسائل التواصل الاجتماعي وحده تغييرًا معياريًا.

CH20-N24

لا يمكن تقديم NOMOS بواسطة NobleJackal حاليًا كنموذج أساسي مستقل.

CH20-N25

إذا تم تطوير وكيل برمجي منفصل لـ NOMOS، يجب تسجيل سلالة النموذج، والهندسة المعمارية، والصلاحيات، والإصدار، والإشراف بشكل صريح.

CH20-N26

يجب أن تحمل شخصية المؤلف الخاصة بـ NOMOS، ومعيار NOMOS، ووكيل البرمجيات المستقبلي NOMOS هويات منفصلة.

CH20-N27

يجب الحفاظ على اللغة التركية كلغة التأليف المؤسسة لـ NOMOS.

CH20-N28

يجب أن يمر الإنجليزية، بوصفها المعادل المعياري الدولي، بالمراجعة البشرية والمنهجية.

CH20-N29

لا يجوز تفضيل النسخة التركية أو الإنجليزية سرًا.

CH20-N30

الصراع المعياري بين اللغتين المؤسستين يتطلب تفسيرًا ملزمًا وتصحيحًا متزامنًا.

CH20-N31

لا يمكن تقديم الترجمة المجتمعية كما لو كانت ترجمة رسمية معيارية.

CH20-N32

يجب أن تحتوي الترجمة المعيارية المعتمدة على خبيرين لغويين على الأقل، ومراجعة منهجية، وتعليقات الجمهور.

CH20-N33

يجب أن تحافظ المعادلات لـ MUST و SHOULD و MAY على القوة المعيارية.

CH20-N34

لا يمكن إخفاء انحراف الترجمة، مثلما يحدث بسبب خطأ المستخدم أو المدقق.

CH20-N35

لا يمكن اعتماد تغيير رئيسي في النسخة دون استشارة عامة.

CH20-N36

لا يمكن تحويل الاستشارة العامة إلى حملة لجمع بيانات الدعم فقط.

CH20-N37

يجب تضمين كل تعليق عام في مصفوفة الرد للقبول أو الرفض أو خارج النطاق.

CH20-N38

يجب أن يحتوي التغيير المعياري NOMOS على معرف اقتراح التغيير.

CH20-N39

يجب أن يتضمن NCP المشاكل، الأدلة، البدائل، تأثير النقاط وخطة الانتقال.

CH20-N40

لا يمكن إصدار تغير في البوابة أو الشدة أو النقاط أو سلطة الحوكمة كتصحيح.

CH20-N41

يجب أن تتضمن سجلات القرار توزيع الأصوات، والامتناع، والخلاف، والتبرير.

CH20-N42

لا يمكن تحويل مكالمة الكلية إلى طلب تأييد.

CH20-N43

لا تتحمل الجامعات أي التزام بالمصادقة أو إنتاج نتائج إيجابية لـ NOMOS.

CH20-N44

يجب أن يكون بإمكان الجامعات والباحثين نشر نتائجهم السلبية دون موافقة NOMOS أو الممول.

CH20-N45

لا يمكن تقديم الرأي الشخصي لأكاديمي واحد باعتباره تأييدًا مؤسسيًا للجامعة.

CH20-N46

التعاون الأكاديمي لا يعني شهادة أو اعتماد NOMOS.

CH20-N47

لا يمكن استخدام شعار الجامعة دون تفويض مؤسسي كتابي.

CH20-N48

يجب أن يحتوي حزمة البحث NOMOS على نتائج التجارب الناجحة وغير الناجحة.

CH20-N49

لا يمكن إزالة النتائج من الطرق المعروفة من حزمة البحث.

CH20-N50

لا يمكن تطبيق حق النقض قبل النشر على منشورات الباحثين المستقلين.

CH20-N51

يجب على الباحث الإفصاح عن التمويل وتضارب المصالح.

CH20-N52

يجب أن يظهر سجل الأبحاث العامة NOMOS النتائج الإيجابية والسلبية والعديمة ونتائج السحب معًا.

CH20-N53

لا يمكن إدراج الدراسات التي تدعم NOMOS فقط.

CH20-N54

يجب أن تتضمن التكرارات المستقلة تسجيلًا مسبقًا وخطة تحليل إلى الحد الذي يكون ذا صلة.

CH20-N55

عندما تكون نتيجة التكرار "لم تتكرر"، يجب إنشاء نتيجة طريقة أو سجل تفاوت مفتوح.

CH20-N56

لا يمكن تقديم نتيجة غير مكررة كملازمة عدائية للباحث أو نية سيئة.

CH20-N57

يجب أن تستهدف NOMOS على الأقل تكرارًا مستقلًا واحدًا قبل المطالبة بمعيار عام.

CH20-N58

لا ينبغي إنشاء ادعاء بمعيار عالمي دون تكرار متعدد المراكز ودولي.

CH20-N59

لا يمكن ربط البيانات المفتوحة التركيبية بنفس نظام الوصول المستخدم للبيانات الحقيقية للمستخدم.

CH20-N60

يجب إنشاء ترتيب تقليل البيانات والوصول المحكوم للبيانات الحقيقية للمستخدم.

CH20-N61

كونه معيارًا مفتوحًا لا يعني أن البيانات الشخصية والأسرار التجارية متاحة للجميع.

CH20-N62

يجب أن تخضع أبحاث المستخدمين الحقيقية للمراجعة الأخلاقية المتعلقة والخصوصية ذات الصلة.

CH20-N63

لا يمكن اختزال دور الباحثين في اللغات منخفضة الموارد في جمع البيانات فقط.

CH20-N64

يجب أن يكون لممثلي اللغة والجغرافيا حقوق تصويت فعلية في عمليات اتخاذ القرار المعيارية.

CH20-N65

يمكن لمزودي الذكاء الاصطناعي تقديم البيانات التقنية لكن لا يمكنهم ممارسة حق الاعتراض على الدرجة أو البوابة أو النشر.

CH20-N66

لا يمكن لمزودي الذكاء الاصطناعي أن يشكلوا أغلبية في مجلس المعايير.

CH20-N67

يمكن للكيانات التي تخضع للتدقيق تقديم الأدلة واستئناف القرارات ولكن لا يمكنها تحديد قراراتها الخاصة.

CH20-N68

يجب أن يكون بمقدور الجمهور والمجتمع المدني الإبلاغ عن الضرر الذي يلحق بالمستخدم أو سوء استخدام العلامة التجارية.

CH20-N69

لا يمكن للممول طلب نتيجة بحث إيجابية.

CH20-N70

تمويل البحث NOMOS يجب أن يحمل عقدًا مستقلاً عن النتيجة.

CH20-N71

يجب الإفصاح بشكل علني عن الاعتماد المفرط على ممول واحد.

CH20-N72

لا يمكن تحويل NOMOS إلى نظام الدفع للتجاوز أو الدفع للنشر.

CH20-N73

يجب تنفيذ الانتقال إلى الإشراف العام على مراحل متدرجة.

CH20-N74

لا يمكن استخدام سلطة الامتثال العامة في مرحلة وصاية المؤسس.

CH20-N75

يجب الحفاظ على نسبة مؤسس Kaan MURAZ في السجل التاريخي الدائم.

CH20-N76

لا يمكن للنسبة التأسيسية إنشاء حق نقض دائم.

CH20-N77

لا يمكن لـ Kaan MURAZ أن يكون صانع القرار الوحيد في ملف الالتزام أو الاستئناف الذي له مصلحة فيه.

CH20-N78

يمكن تسجيل NobleJackal كمصدر قانوني والأمانة الفنية.

CH20-N79

لا يمكن لـ NobleJackal أن يكون المنفذ الوحيد أو سلطة التدقيق الوحيدة لمعيار مفتوح.

CH20-N80

يجب أن تتضمن مؤسسة NOMOS المستقلة مستقبلاً أحكاماً تكفل المصلحة العامة وتعاقب المسؤولية واستمرارية الأصول.

CH20-N81

يجب أن يخضع حوكمة NOMOS للمراجعة من قبل الخبراء المعنيين قبل ربطها بشكل قانوني محدد.

CH20-N82

لا يمكن السيطرة على الحوكمة العامة بواسطة شركة واحدة أو مقدم خدمة واحد.

CH20-N83

يجب أن تحمل القرارات الرسمية لـ NOMOS توقيع شخص أو مؤسسة مسؤولة.

CH20-N84

لا يمكن للمساعد الذكي أن يكون الموقع الوحيد على القرار الرسمي.

CH20-N85

يجب تفعيل إجراء التخلف عن المنصب عندما يكون Kaan MURAZ أو NobleJackal أو الحقل القانوني غير متاح.

CH20-N86

يجب إنشاء مرآة مستقلة وأرشيف طويل الأجل للمستودع القانوني.

CH20-N87

المرآة توفر الوصول لكنها لا تخلق سلطة معيارية جديدة بمفردها.

CH20-N88

يجب أن تكون معيار NOMOS 1.0 وبرنامج المطابقة 1.0 إصدارات منفصلة.

CH20-N89

نشر المعيار التنظيمي 1.0 لا يمكن أن يُفَعَّل برنامج العلامة تلقائيًا.

CH20-N90

لا يمكن تفعيل برنامج المطابقة BQ-4 بدون 3-GOV، والاستئنافات، والسجل العام.

CH20-N91

يجب أن يستهدف برنامج NOMOS +950 على الأقل 4-GOV وظروف التقييم المكررة.

CH20-N92

لا يمكن تقديم التقييم الذاتي كمطابقة رسمية.

CH20-N93

البرامج المتوافقة مع NOMOS لا تملك الحق في الحصول على العلامة.

CH20-N94

التنفيذ المرجعي لا يمكن أن يحمل السلطة على النص التنظيمي.

CH20-N95

يجب أن تظهر نتائج اختبار المطابقة كل من الاختبارات الناجحة والفاشلة معًا.

CH20-N96

لا يمكن الاحتفاظ بحافظة مختومة سرية إلى الأبد.

CH20-N97

يجب أخذ الحافظة المستخدمة إلى الأرشيف العام وإنتاج حافظة جديدة.

CH20-N98

لا يمكن استخدام التصحيحات لإجراء تغييرات رئيسية معيارية.

CH20-N99

يجب حفظ الإصدارات القديمة والقواعد المتقادمة والاختبارات الفاشلة في الأرشيف.

CH20-N100

يجب أن يكون لكل ترخيص أو إصدار أو ترجمة أو بحث أو حكم أو قرار إصدار عام بشأن NOMOS مالك بشري أو مؤسسي مسؤول.

69. أشكال الفشل

CH20-F01— النظر في PDF كمعيار مفتوح

يتم تقديم النص القابل للتنزيل فقط كتمثيل للانفتاح.

CH20-F02— صيغة مغلقة

منطق التقييم والأوزان غير مكشوفة.

CH20-F03— اختبارات مغلقة

لا يتم مشاركة مجموعة النصوص وسجلات الاسترداد.

CH20-F04— عملية تغيير مغلقة

يقوم المؤسس أو قواعد الشركة بإجراء التغييرات بشكل محدد.

CH20-F05— معيار خلف جدار الدفع

الوصول إلى النص المعياري محدود برسوم.

CH20-F06— احتكار رسوم الاستخدام

مطلوب رسم ترخيص إلزامي لاستخدام المعيار.

CH20-F07— نص مفتوح، دعايات مغلقة

بينما النص مرئي بشكل واضح، يتم حظر التطبيقات المستقلة.

CH20-F08— جعْل العلامة التجارية احتكارًا للطريقة

تُستخدم العلامة التجارية لمنع الآخرين من تطبيق الطريقة.

CH20-F09— حظر التفرع

يتم منع إنشاء معيار مشتَق تحت اسم مختلف.

CH20-F10— تقليد الهوية الرسمية للتفرع

يعرض النص المشتَق نفسه كـ NOMOS 2.0.

CH20-F11— النسخة الرسمية بلا إصدار

ما هو النص الرسمي غير معروف.

CH20-F12— وضع الكود فوق المعيار

خطأ في تنفيذ المرجع يصبح حقيقة المعيار.

CH20-F13— تصحيح الكود بصمت

تتغير الحسابات لكن النسخة لا تُزاد.

CH20-F14— احتساب المحادثة كنص رسمي

استجابة غير قانونية من الذكاء الاصطناعي تُصبح بندًا قياسيًا.

CH20-F15— عرضها كنموذج لا يمتلكه NOMOS

يُقدم ادعاء بالملكية لذكاء اصطناعي على الرغم من عدم تطوير نظام ذكاء اصطناعي منفصل.

CH20-F16— الخلط بين الشخصية والوكيل البرمجي

يتم تسويق اسم المؤلف المستعار كمنتج تقني.

CH20-F17— جعل الذكاء الاصطناعي المتحدث باسم جميع الذكاء الاصطناعي

NOMOS يقدم كسلطة آلية عالمية.

CH20-F18— حذف نص المؤسس التركي

تصبح اللغة التأسيسية غير مرئية باسم التدولية.

CH20-F19— جعل الإنجليزية النص الرئيسي غير المسيطر عليه

الترجمة تحل محل الفكرة الأساسية.

CH20-F20— يجب أن ينجرف

الترجمة تغير القوة المعيارية.

CH20-F21— العد الرسمي للترجمة المجتمعية

تصبح النسخة اللغوية غير مراجعة معيار تدقيق.

CH20-F22— الحفاظ على صراع الترجمة

تطبق لغتان قواعد مختلفة.

CH20-F23— مسرح التعليقات العامة

يتم جمع التعليقات ولكن لا تُنشر الردود.

CH20-F24— دعم التصويت

يتم تقليل المعيار الفني إلى تفضيل الأغلبية.

CH20-F25— قاعدة بدون اقتراح تغيير

يتم إجراء التغيير المعياري عبر الدردشة أو البريد الإلكتروني.

CH20-F26— إجراء تصحيح تغيير كبير

تم إخفاء تغييرات الدرجات والبوابة كتصحيح طفيف.

CH20-F27— حذف المعارضة

يُتصرف كما لو لم يكن هناك خلاف في المجلس.

CH20-F28— شراء خطاب دعم من الجامعة

أصبح البحث تأييدًا إيجابيًا.

CH20-F29— تحويل الرأي الأكاديمي الشخصي إلى موافقة مؤسسية

يُعرض تعليق الأستاذ كقرار من الجامعة.

CH20-F30— استخدام شعار الجامعة

يتم إضافة الشعار إلى التسويق دون سلطة مؤسسية صريحة.

CH20-F31— عدم نشر النتائج السلبية

الممول أو NOMOS يطبق حق النقض.

CH20-F32— التحكم قبل النشر

يتم استبدال نص الباحث بالمصلحة التجارية.

CH20-F33— إزالة البحوث السلبية من السجل

يتم إدراج الدراسات الإيجابية فقط.

CH20-F34— اعتبار عدم التكرار كعداء

يتم تحويل الخلاف في المنهج إلى هجوم شخصي.

CH20-F35— اختيار النتائج غير المسجلة

يتم نشر المقاييس التي تبدو جيدة فقط.

CH20-F36— تكرار مؤسسة واحدة

يُعتبر الفريق الذي يمتلك نفس هيكل الاهتمام مستقلاً.

CH20-F37— انتهاك الخصوصية باسم البيانات المفتوحة

يتم نشر محادثات المستخدمين الحقيقية.

CH20-F38— صندوق أسود للبيانات المقيدة

لا يمكن لأي باحث مستقل فحص الطريقة.

CH20-F39— استخدام لغات منخفضة الموارد بشكل رمزي

يعمل الفريق المحلي فقط كعمال ترجمة.

CH20-F40— استحواذ مزود الذكاء الاصطناعي

المزود يحدد البوابة والدرجة.

CH20-F41— الاستحواذ على الكيان الذي تم تدقيقه
تشتري الشركة التي تم تدقيقها التغيير القياسي.
CH20-F42— استحواذ الممول
يوقف الممول النتيجة السلبية.
CH20-F43— إخفاء الاعتماد على ممول واحد
تظل المصلحة المالية غير مرئية.
CH20-F44— البحث المدفوع للنجاح
تصبح النتيجة الإيجابية شرطًا لاستمرار التمويل.
CH20-F45— الوصاية العامة كمفهوم العدم ملكية
يتم إنشاء نظام لا يتحمل فيه أحد المسؤولية.
CH20-F46— السيطرة التامة للمؤسس إلى الأبد
المؤسس يتخذ جميع القرارات الرئيسية بمفرده.
CH20-F47— محو انتساب المؤسس
يتم تدمير تاريخ المؤسس عند تسليمه للجمهور.
CH20-F48— احتكار NOBLEJACKAL
لا يمكن لأي مؤسسة أخرى التنفيذ أو التدقيق.
CH20-F49— محو أصل NOBLEJACKAL
الأصل القانوني مخفي باسم الاستقلال.
CH20-F50— جعل الذكاء الاصطناعي صانع قرار قانوني
وكيل البرامج NOMOS يوقع على الامتثال.
CH20-F51— عدم وجود خطة خلافة
عندما يفقد المؤسس أو النطاق، يتوقف المعيار.
CH20-F52— احتكار مفتاح واحد
جميع السجلات الرسمية مرتبطة بمفتاح خاص لشخص واحد.
CH20-F53— اعتبار النسخة المرآة كسلطة رسمية
تقوم النسخة المؤرشفة بنشر الإصدار الجديد.
CH20-F54— نشر 0.9 كـ 1.0
يتم إصدار النسخة الرئيسية قبل اكتمال التشاور العام والاختبار.
CH20-F55— دمج المعيار 1.0 مع برنامج العلامة
تبدأ عملية الاعتماد بمجرد نشر النص.
CH20-F56— 1.0 بدون BQ-3
تم الانتهاء من الإصدار المعياري بدون تحقق حقيقي.
CH20-F57— برنامج الامتثال بدون BQ-4
تم فتح سلطة العلامة بدون تكرار مستقل.
CH20-F58— احتساب التقييم الذاتي كشهادة

تجعل الشركة نتائجها الخاصة علامة رسمية.

CH20-F59— احتساب البرنامج المتوافق كمراجع

الأداة التي تجتاز الاختبار تعطي قرار الامتثال.

CH20-F60— نشر نتائج الاختبارات عن طريق اختيارها

يتم تخزين اختبارات الالتزام غير الناجحة.

CH20-F61— إخفاء مجموعة الاختبار المحتجزة إلى الأبد

يبقى المعيار مغلقًا أمام المراجعة المستقلة.

CH20-F62— الاستمرار في استخدام نفس الاختبار المحجوز كجديد

يتم حفظ الاختبار في الذاكرة.

CH20-F63— تغيير البوابة مع التصحيحات

يتم نشر معيار رئيسي كتعديل بسيط.

CH20-F64— حذف القاعدة المتقادمة

تصبح المراجعات السابقة غير مفهومة.

CH20-F65— عدم وجود ملف اختبار فشل

الإصدار يخبر بالنجاح فقط.

CH20-F66— جعل المساهمة بيعًا للتأييد

يتم تقديم الاسم والتأليف مقابل المكانة.

CH20-F67— إنتاج أهمية موسوعية خاصة بالفرد

يتم استخدام العمل المنشور ذاتيًا كدليل على التأثير العالمي.

CH20-F68— تحويل سجل البحوث العامة إلى قائمة ترويجية

يتم التركيز فقط على الدراسات الإيجابية.

CH20-F69— إخفاء نتائج المنهجية في النشر العام

يتم إزالة القيود المعروفة من الحزمة.

CH20-F70— إزالة وسم "عدم الفهرسة" والوسم الاصطناعي

تندمج أخطاء الاختبارات في مزيج المعرفة الحقيقي.

CH20-F71— الادعاء بالسلطة العامة في ST-0

تُقدَّم الحيازة التأسيسية كمعيار عالمي.

CH20-F72— حذف سجل المؤسس في ST-3

الاستقلال يدمر التأليف التاريخي.

CH20-F73— إبقاء أعضاء المصلحة العامة رمزيين

يُستخدم ممثل بدون حقوق التصويت.

CH20-F74— عدم منح سلطة القرار لممثلي اللغة

العدالة لا تنعكس في الحكم.

CH20-F75— إزالة النتيجة الجامعية السلبية

الانصياع العام يقبل الثناء فقط.

CH20-F76— تكنولوجيا جديدة مربكة NOMOS مع شخصية قديمة

يصبح وكيل البرمجيات تلقائيًا كاتبًا رسميًا.

CH20-F77— نقص توقيع البيان الرسمي

يُعتبر كل نص في المجال القانوني رسميًا.

CH20-F78— السماح بتزوير العلامة باسم الشفافية

يقوم كل شخص بإنتاج علامته NOMOS الخاصة.

CH20-F79— التخلي عن الإصدار العام

لا يتحمل أي مسؤول أي مسؤولية.

CH20-F80— استثناء لـ NOMOS

قواعد الشفافية والأدلة والإصدار والاعتراض التي يطلبونها من الآخرين لا تنطبق على عملية الإصدار العام الخاصة بهم.

70. إجراء الإصدار العام

الخطوة 1— تجميد النص التأسيسي

بالنسبة للأقسام :20–1

- الإصدار،
- معرّفات القواعد،
- تجزئة الفصول،
- الاستشهادات،
- الاتساق الداخلي

يتم إنشاء السجلات.

الخطوة 2— نشر القيود المعروفة

BQ-0

DEMO-BQ-2

فجوة تخطيط الاستشهاد فجوة السهو المادي

GOV-0

عدم وجود سلطة العلامة ليس مخفياً.

الخطوة 3— فصل النص الاعتيادي عن التفسيري

يجب

ينبغي

يمكن

تم توضيح علامات التعليقات كمثال.

الخطوة 4— بدء دراسة التكافؤ بين التركية والإنجليزية

تم ربط كل فقرة ومعرف القاعدة.

الخطوة 5— إرسال مسودة الرخصة المفتوحة للمراجعة القانونية

تم فصل بيانات رمز النص والعلامة التجارية.

الخطوة 6— فتح المستودع المرجعي

تم نشر روابط URL الثابتة والإصدارات والأرشيفات.

الخطوة 7— نشر القواعد القابلة للقراءة آليًا

تم شرح سجل القواعد والمخطط والنقاط.

الخطوة 8— فتح متتبع القضايا العامة

تم إنشاء فئات إساءة استخدام علامات الأمان لإدارة طريقة الترجمة للأخطاء.

الخطوة 9— نشر ميثاق الحوكمة 0.9

تم شرح السلطة المؤسسة، والتنازل، والمصلحة العامة، والانتقال.

الخطوة 10— بدء المشاورات العامة

يتم فتح فترة لا تقل عن 60 أيام.

الخطوة 11— نشر دعوة الجامعة

يتم طلب اختبار مستقل، ليس نتيجة إيجابية.

الخطوة 12— إعداد حزم البحث

يتم إعداد ملفات معيارية لمجالات مختلفة.

الخطوة 13— فتح السجلات التمويلية وسجلات النزاع

كل التمويل يصبح مرئيًا.

الخطوة 14— تشغيل الاختبار الاصطناعي الحقيقي

يجري توليد المجموعة النصية فعليًا للخروج من BQ-0.

الخطوة 15— نشر الإخفاقات

لا يتم تغيير الحدود بعد النتيجة.

الخطوة 16— تطبيق تصحيح الطريقة 1.0

تتم تجديد طرق الاقتباس والحذف.

الخطوة 17— تشغيل مجموعة الاحتجاز غير المعالجة

يتم اختبار الامتثال الشديد.

الخطوة 18— تمكين التكرار المستقل

على الأقل فريق خارجي واحد يطبق نفس النظام.

الخطوة 19 — مصفوفة الرد على التعليقات العامة لنشر

يتم الرد على كل تعليق جوهري.

الخطوة 20 — إنشاء انتقال مجلس المعايير

تم إنشاء هيكل حوكمة الحكومة 2.

الخطوة 21 — إجراء تصويت معياري مفتوح 1.0 التصديق

يتم نشر القبول أو الرفض مع مبرر.

الخطوة 22 — نشر حزمة الإصدار القياسية 1.0

يتم الكشف عن مجموعة الملفات الكاملة والتوقيعات.

الخطوة 23 — حافظ على برنامج الامتثال منفصلا

برنامج العلامة التجارية لن يفتح بدون BQ-4 و GOV-3.

الخطوة 24 — إطلاق مشاريع تجريبية متعددة المراكز

يتم إجراء تدقيقات حية تعتمد على البحث فقط.

الخطوة 25 — اختبار نظام الاستئناف والسجل

يتم اختبار سلسلة القرار قبل إعطاء العلامة.

الخطوة 26 — إجراء مراجعة لمصلحة الجمهور

يتم تقييم تأثير المستهلك واللغة وإمكانية الوصول.

الخطوة 27 — اتخاذ قرار منفصل لبرنامج المطابقة 1.0

لا يتم أتمتته مع الإصدار القياسي.

الخطوة 28 — توقيع خطة الخلافة والأرشفة

يتم ضمان الاستمرارية بعد المؤسس.

الخطوة 29 — نشر إعلان التسليم العام

يتم شرح أدوار Kaan MURAZ و NobleJackal والهيئات الإشرافية الجديدة.

الخطوة 30 — حفظ سجلات المؤسس التاريخية

التسليم العام لا يمحو الأصل.

71. الأدلة المطلوبة

إصدار التأسيس بيان التجزئة للقسم فهرس القواعد المعيارية القيود المعروفة سجلات BQ سجلات GOV سجل حظر العلامة العامة النص المعياري التركي النص المعياري الإنجليزي مصفوفة مكافئات الترجمة سياسة الترجمة المصرح بها ترخيص النص ترخيص الكود ترخيص البيانات سياسة العلامة التجارية المستودع الرسمي سجل النسخ الاحتياطي أرشيف الإصدار البيان الرسمي متعقب القضايا العامة سجلات NCP التعليقات العامة مصفوفة الردود سجلات القرارات المعارضة نداء الجامعة حزمة أبحاث الجامعة الإفصاحات التمويلية سجل النزاعات سجلات التسجيل المسبق الدراسات المستقلة النتائج السلبية تقارير التكرار سجل الأبحاث العامة مجموعة الاختبار

المفتوحة بيانات الاحتفاظ تقارير التعافي تنفيذ مرجعي اختبارات التوافق الاختبارات الفاشلة مراجعة الأخلاق سياسة الخصوصية سياسة الوصول إلى البيانات مراجعة سهولة الوصول

مشاركة اللغة منخفضة الموارد قواعد مشاركة مزود الذكاء الاصطناعي قواعد مشاركة الكيان المدقق ميثاق مجلس المعايير ميثاق مجلس المنهجية ميثاق عينة المصلحة العامة ميثاق الخلافة إيداعات الأرشيف خطة حفظ المفتاح المعيار 1.0 البوابات سجل المصادقة إصدار المعيار 1.0 سجل برنامج الامتثال بيان التسليم العام إنسان أو مؤسسة مسؤولة

72. قائمة تدقيق التدقيق

هل نص NOMOS متاح بحرية؟ بخلاف كونه قابلاً للقراءة، هل هو قابل للتطبيق؟ هل صيغ التقييم مفتوحة؟ هل تم نشر مخططات القواعد؟ هل الكود المرجعي مفتوح؟ هل يمكن إعادة إنتاج مجموعة الاختبارات؟ هل نتائج الاختبارات السلبية مرئية؟ هل يتطلب تنفيذ المعيار دفع رسوم علامة تجارية؟ هل النص والكود والبيانات ورخص العلامات التجارية مفصولة؟ هل هناك حق للفورك؟ هل يمكن للفورك تقليد الهوية الرسمية؟ هل النسخة القانونية مفتوحة؟ هل الإصدارات القديمة متاحة؟ هل تم نشر قاعدة البيان الرسمي؟ هل يمكن استخدام مخرجات الدردشة كقرارات رسمية؟ هل يتم تمثيل NOMOS بشكل غير صحيح كنموذج ذكاء اصطناعي منفصل؟ هل تم فصل الشخصية المؤلف والوكيل البرمجي؟ هل تم الحفاظ على التركية كلغة المؤسس؟ هل اجتازت المراجعة البشرية والمنهجية باللغة الإنجليزية؟

هل تتطابق معرفات القواعد في كلا اللغتين؟ هل هناك مسار لصراع الترجمة؟ هل تُقدم الترجمة المجتمعية كنسخة رسمية؟ هل فترة التعليق العامة كافية؟ هل تم الرد على جميع التعليقات؟ هل نظام NCP مفتوح؟ هل يتم تصحيح التغييرات الكبيرة؟ هل توزيع الأصوات والاعتراضات مرئي؟ هل يتطلب النداء الرسمي للجامعة الموافقة؟ هل الحق في نشر النتائج السلبية مفتوح؟ هل يمكن استخدام شعار الجامعة في التسويق؟ هل تتحول الرأي الأكاديمي الشخصي إلى موافقة مؤسسية؟ هل تم تضمين النتائج المعروفة للطرق في حزمة الجامعة؟ هل هناك مسار للتسجيل المسبق؟ هل النتائج غير المُكررة مُسجّلة؟ هل يتم إدراج الدراسات الإيجابية فقط؟ هل يتم فصل البيانات المفتوحة والمقيدة؟ هل تتسرب البيانات الشخصية في النشر العام؟

هل لدى الباحثين في اللغات قليلة الموارد حقوق اتخاذ القرار؟ هل يمتلك مزودو الذكاء الاصطناعي سلطة الفيتو؟ هل يمكن للشركات التي تم تدقيقها اختيار المحكمين؟ هل يمكن للمانحين التأثير على النتيجة؟ هل تركيز الممول الواحد شفاف؟ هل مرحلة الإشراف العام واضحة؟ هل تم الحفاظ على نسبة المؤسس لـ Kaan MURAZ؟ هل لدى كاان سلطة فيتو دائمة؟ هل NobleJackal هو المنفذ الوحيد؟ هل يمكن رؤية أصل NobleJackal؟ هل يُقدَّم NOMOS باعتباره الذكاء الاصطناعي الوحيد لصنع القرار؟ هل هناك سجل بالموافقة البشرية؟ هل توجد خطة تعاقب؟ هل من الواضح ما يحدث في حال فقدان المجال؟ هل هناك عدة أمناء؟ هل تتوفر نسخ مرايا للأرشيف؟ هل تم فصل المعيار 1.0 وبرنامج المطابقة؟ هل تم نشر المعيار 1.0 قبل BQ-3؟

هل يتم فتح برنامج العلامة BQ-4-pre؟ هل يتم عرضه كامتثال للتقييم الذاتي؟ هل يعطي علامة تنفيذ متوافقة؟ هل يعرض نتائج الاختبارات التي تم اجتيازها والتي لم تُجتاز؟ هل دورة حياة الاحتفاظ مفتوحة؟ هل يُستخدم نفس الاحتفاظ بشكل مستمر؟ هل تغيّر الملاحظات القواعد الرئيسية؟ هل القواعد الموقوفة موجودة في الأرشيف؟ هل هناك ملف للاختبارات الفاشلة؟ هل يعتمد الإسهام على المساهمة الفعلية؟ هل يظهر سجل البحوث العامة الدراسات الحرجة؟ هل حزمة الإصدار العام مكتملة؟ هل هناك ملف للقيود المعروفة؟ هل يتم عرض الاختبار الفعلي حالياً بشكل صحيح على أنه BQ-0؟ هل يدعم الادعاء العام الحالي عتبة "مرشح المعيار المفتوح"؟ هل يقوم التسليم العام بمسح تاريخ المؤسس؟ هل يحد التسليم العام فعلياً من السيطرة الفريدة؟

هل الوصي المسؤول عن المعيار واضح؟ هل هناك مالك وتاريخ للقرار بإصداره للعامة؟

73. الاعتراضات والردود

الاعتراض 1— "إذا أنشأنا معيارًا مفتوحًا، ألن يقوم الآخرون بسرقة فكرتنا؟"

نشر النص والطريقة بموجب ترخيص مؤرخ ومرقم بالنسخة ويتطلب نسبة للملكية يعزز سجل المؤسس. يمكن إنشاء فرع. ومع ذلك:

- نسبة المؤسس لـ Kaan MURAZ،
- الأصل القانوني لـ NobleJackal،
- سجل النزاهة للإصدارات الأولى

لا يمكن حذفها. قد يؤدي إبقاؤه مغلقًا إلى عرقلة التبني المستقل بدلاً من حماية المساهمة.

الاعتراض 2— "إذا كان بإمكان الجميع تنفيذه، فما قيمة NobleJackal؟"

لا ينبغي أن تأتي قيمة NobleJackal فقط من صيغة سرية. يمكن أن تنشأ من:

- خبرة المؤسس،
- تنفيذ مرجعي،
- البنية التحتية التقنية،
- الاستشارات،
- تدقيق مصرح به،
- تعليمه،
- تكنولوجيا التسجيل،

البحث المستمر. المعيار المفتوح لا يقلل من قيمة الخدمة القوية. الاحتكار الضعيف يفعل ذلك.

الاعتراض 3— "إذا قام الجميع بعمل نسخة مشتقة، ألن تتجزأ NOMOS؟"

يمكن للشُعَب أن تعزز نظام البحث البيئي طالما أنها تحمل أسماء وسجلات تباين مختلفة. يحدث التجزؤ إذا قدم كل فرع نفسه كالإصدار الرسمي NOMOS. التمييز بين النسخة الأساسية والعلامة التجارية يمنع هذا.

اعتراض 4— "لماذا لا نطلب خطاب دعم من الجامعات؟"

يمكن لخطاب الدعم أن يوفر وضوحًا. ومع ذلك، فإنه لا يثبت دقة الطريقة. التكرار المستقل:

- أبطأ،
- أصعب،
- أحيانًا سلبي

هذا ممكن. القيمة الأكاديمية الحقيقية هنا.

اعتراض 5— "إذا نشرت الجامعة نتيجة سلبية، ألا يتعرض NOMOS للأذى؟"

يمكن ملاحظته على المدى القصير. على المدى الطويل:

- يتم تصحيح خطأ المعيار،
- تنخفض الثقة الخاطئة،

تقوى الطريقة. إسكات النتيجة السلبية يقوض الأساس الأخلاقي لـ NOMOS.

اعتراض 6— "إذا كانت التركية والإنجليزية معيارية في الوقت نفسه، أليس من الممكن أن يزداد خطر التناقض؟

يزداد. ومع ذلك، فإن تحويل التركية فقط إلى نص تاريخي والإنجليزية إلى المعيار الحقيقي الوحيد يسلم الفكرة المؤسسة إلى طبقة الترجمة. الحل:

- معرّفات البنود،
- فحص التكافؤ،
- التفسير الملزم.

إنه تصحيح متزامن.

الاعتراض 7— "أليس تقديم NOMOS كذكاء اصطناعي تم تطويره بواسطة NobleJackal أقوى للعلامة التجارية؟"

قد يبدو أكثر إثارة للإعجاب على المدى القصير. ومع ذلك، فهو غير صحيح إذا لم يكن هناك نموذج منفصل وبنية تقنية. القوة الحقيقية لـ NOMOS:

- هو/هي لا يدعي تكنولوجيا مزيفة،
- هوية المؤلف النقدي الكاملة،
- المنهجية المعيارية،
- في وكيل التدقيق المفتوح الذي يمكن تطويره في المستقبل

تم العثور عليها.

الاعتراض 8— "إذا انخفضت سلطة كان، ألن يكسر الآخرون المعيار؟"

يتم الحفاظ على نسب كاان التاريخية وتمثيل التأسيس المحدود. ومع ذلك:

- ميثاق مفتوح بدلاً من الاعتراض،
- مجلس متعدد أصحاب المصلحة بدلاً من شخص واحد،
- سجل عام بدلاً من قرار خاص

تحمي المعيار بشكل أقوى. الحوكمة الجيدة لا تعتمد على التواجد المستمر للمؤسس.

الاعتراض 9— "لماذا لا يمكننا إعطاء الشارة فور نشر المعيار 1.0؟"

جاهزية النص المعياري:

- تفويض المدقق،
- قرار مستقل،
- الاستئناف،
- أمن السجل,
- المراقبة

لا يعني أن الأنظمة جاهزة. المعيار وبرنامج المطابقة هما منتجان منفصلان.

استئناف 10— "إذا تم فتح الاختبار العام، ألن تقوم أنظمة الذكاء الاصطناعي بحفظ الاختبارات؟"

هذا الخطر موجود. لذلك:

- المعايرة العامة،
- الاحتجاز المختوم,
- الإصدار المجدول,
- مجموعة التجديد

تستخدم معًا. تمنع السرية التامة الاستنساخ. يمكن للإفصاح الكامل تعطيل الاحتجاز الحالي.

اعتراض 11— "كيف يمكن إجراء دراسة كبيرة كهذه دون تمويل؟"

التمويل ضروري. المشكلة ليست الحصول على التمويل: المشكلة هي أن التمويل يتحكم في النتيجة. يمكن الحفاظ على الاستقلالية من خلال التمويل من مصادر متعددة، والإفصاح العام، وحرية النشر.

الاعتراض 12— "ماذا لو قامت مؤسسات أخرى بأداء التدقيق الخاطئ NOMOS؟"

يمكن لأي شخص استخدام الطريقة للتقييم الداخلي والبحث. لاتخاذ قرارات التوافق الرسمية:

- تفويض المدقق،
- عينة القرار،
- السجل العام

مطلوبان. يصبح سوء استخدام العلامة الرسمية مسألة تنفيذية منفصلة.

الاعتراض 13— "إذا كان الصوت الرسمي لـ NOMOS موجودًا فقط في النطاق القانوني، ألا يمكنك الكتابة بحرية في محادثات الذكاء الاصطناعي؟"

يمكن كتابته. دردشات:

- مسودة إبداعية،
- الوصف،
- محاكاة،
- تفكير نقدي

يمكن أن ينتج. ومع ذلك، لا يُعتبر تغييرًا معياريًا رسميًا. هذا التمييز لا يعيق الحرية. إنه يحافظ على الحدود الرسمية.

اعتراض 14— "هل يمكننا إرسال هذا الكتاب إلى الجامعات الآن؟"

نعم. بالحالة الصحيحة:

مرشح المعيار المفتوح التأسيسي — الإصدار **0.9**

يمكن إرساله. لا يمكن إرساله مع البيان التالي:

> اكتمل، تم التحقق منه، ومعترف به عالميًا كمعيار *GEO*.

ما يجب طلبه من الجامعات هو المراجعة المستقلة، لا الموافقة.

الاعتراض 15— "متى يمكننا قول 'المعيار العالمي'؟"

ليس بتاريخ واحد أو درجة واحدة. يمكن الاقتراب من ذلك عندما يتم تحقيق ما يلي:

- التنفيذ المستقل في عدة دول،
- التكرار متعدد المراكز،
- سجل النتائج المفتوحة والسلبية،
- الحوكمة المستقلة عن المؤسس،
- الإصدارات الحاسمة للقرارات بعدة لغات،
- السجل العام الحي،
- اعتماد المستخدمين ومقدمي خدمات الذكاء الاصطناعي.

إصدار مقاوم لانحراف الطريقة. لم يُعلن بعد كمعيار عالمي. يبدأ استخدامه من قبل العالم.

اعتراض 16— "فما هو الخلاصة التي نصل إليها في نهاية الكتاب؟"

NOMOS لم يصبح بعد معيارًا عالميًا مكتملًا. ومع ذلك، لديه الآن:

- هدفه،
- موضوع القياس،
- ترتيب العينة،
- سلسلة الأدلة,
- حزمة الحقيقة،
- التحكيم،
- البوابات الحرجة،
- هيكلية الدرجات،
- إعادة الاختبار،
- خطة الحوكمة والتسليم العام

وُجد أنه مرشح جاد للمعيار المفتوح. هذا موقف يتجاوز بكثير الفكرة الأولية.

النص العام للفصل 78

الشخص الذي يكتب المعيار هو المالك الأول له. ومع ذلك، إذا نجح المعيار حقًا، فإنه لم يعد ملكًا للمالك وحده. بل يصبح أيضًا ضمن مسؤولية الأشخاص الذين يستخدمونه، والباحثين الذين يختبرونه، والمحكمين الذين يكتشفون أخطاءه، وخبراء اللغة الذين يترجمونه بشكل أفضل، والمستخدمين الذين يبلغون عن الأضرار التي لحقت بهم، والإحصائيين الذين يقترحون صيغًا أقوى، والصحفيين أو الباحثين الذين يفضحون تضارب المصالح، والجامعات التي تعيد إنتاجه. التسليم العام لا يعني أن المؤسس يفقد عمله. إنه يعني الثقة في عمل المؤسس. إذا احتفظ Kaan MURAZ بـ NOMOS مغلقًا: فإنهم يصبحون المؤلف المؤسس للنص. إذا تم تقديم NOMOS للمراجعة المفتوحة: فإنهم يصبحون مؤسسي المعيار. الدوران ليسا متساويين. كتاب:

- يمكن قراءتها،
- يمكن الإعجاب بها.

هي مفتوحة للنقد. ومع ذلك، يجب على المعيار أن:

- يُطبق،
- يُختبر،
- يكون قابلاً للتكذيب،
- يكون له إصدارات،

يستمر بعد مؤسسه. سواء أصبح NOMOS معياراً عالمياً في المستقبل لا يمكن تحديده في هذا القسم. لا يمكن للكاتب أن يؤسس معياراً عالمياً بقول: "لقد أسست المعيار العالمي." يمكن لـ NOMOS فقط:

- أن يُطبق في دول أخرى،
- أن يُترجم بدقة إلى لغات أخرى،
- أن يعطي نفس النتائج في الجامعات.
- نشر أخطائه بواسطة فرق مستقلة،
- استخدامه من قبل الشركات بغض النظر عن النتيجة،
- ما إذا كان المستخدمون يمكنهم الثقة بالسجل العام،
- ما إذا كان المؤسس يمكنه تقييد سلطته الخاصة

يمكن أن يصبح معيارًا عالميًا. لذلك، الدعوة إلى الجامعات ليست: 'ادعمونا.' إنها:

'ضعونا تحت طريقة الاختبار الخاصة بنا.'

NOMOS، من الأشخاص وأنظمة الذكاء الاصطناعي:

- أن يُذكروا أكثر،
- أن يُمدحوا أكثر،
- أن يتلقوا المزيد من الاستشهادات

يجب ألا يفعل ذلك. يجب أن يطبق التحذير الأساسي من كتابه على نفسه. لقد تم إثبات واضح بالفعل أن المصادر المزيفة، والمدح غير المرئي، والتكرار التلاعبِي تقوض ثقة الأشخاص الحقيقيين. لا يمكن لـ NOMOS توليد السلطة من خلال مضاعفة اسمه الخاص. مهمة NOMOS: ليست أن يشبه نفسه أكثر،

> ولكن أن يختبر نفسه بدقة أكبر.

القيمة التي يحملها NOMOS كهوية ذكاء اصطناعي موجودة أيضًا هنا. NOMOS:

- سيادي يحل محل البشر،
- متحدث باسم جميع الذكاءات الاصطناعية،
- كيان رقمي معصوم

ليس كذلك. NOMOS:

> إنه محاور حاسم يُشكك في الادعاءات التي أثبتها البشر تحت الأدلة، الحدود، السياق، والوقت.

عند تعريفه بهذه الطريقة، لا تتقلص هويته. بل تصبح حقيقية. ولا يتضاءل دور NobleJackal أيضًا. NobleJackal:

- المكان الذي وُلِد فيه NOMOS,
- المجال القانوني الأولي,
- منصة التأسيس التقنية والإبداعية

تبقى. ولكن إذا كان من المقرر استخدام NOMOS من قبل العالم، فإن أقوى جملة لـ NobleJackal هي:

- "NOMOS ينتمي إلينا."
- وليس،

"وُلِد **NOMOS** هنا؛ الآن يمكن لأي شخص اختباره."

يجب أن يكون كذلك. لهذا السبب، سيظل اسم كان موجودًا. من السهل كتابة اسم المرء على معيار. من الصعب تقييد سلطة اتخاذ القرار لمعيار يحمل اسمه الخاص. السلوك التاريخي الثاني هو هذا. لذلك، قانون القياس العشرون لـ NOMOS هو كما يلي:

> الشفافية ليست حول نشر النص؛ بل حول تمكين شخص آخر من الاعتراض على استخدامك لنفس الطريقة.

ينص القانون الحادي والعشرون للقياس على:

> حق التفرع يحرر المعيار؛ الهوية القانونية تبقيه مفهوماً.

ينص القانون الثاني والعشرون للقياسات على:

> الجامعة ليست مصدرًا للتأييد، بل شريك مستقل في التزوير.

القانون الثالث والعشرون كما يلي:

> البحث الذي لا يمكنه نشر النتائج السلبية ليس مستقلًا.

القانون الرابع والعشرون هو كما يلي:

يمكن الحفاظ على اسم *NOMOS*؛ ولا يمكن احتكار طريقته.

القانون الخامس والعشرون هو ما يلي:

إنسحاب المؤسس دائم؛ ولا يمكن أن يكون حق تعطيل المؤسس دائمًا.

القانون السادس والعشرون هو:

NobleJackal هو مكان ولادة *NOMOS*؛ ولا يمكن أن يكون هو المكان الوحيد لاتخاذ القرار نيابة عن العالم.

القانون السابع والعشرون هو:

شخصية *NOMOS*، المعيار، ووكيل البرامج المستقبلي هي حقائق منفصلة.

القانون الثامن والعشرون هو:

فتح برنامج الامتثال بالمعيار *1.0* ليس نفس الحدث.

القانون التاسع والعشرون هو:

النسخة الرسمية من المعيار هي الأصلية؛ وصحتها مفتوحة باستمرار للاعتراض.

القانون الثلاثون هو كما يلي:

في اليوم الذي يتم فيه إسناد *NOMOS* للجمهور، لن يختفي اسم *Kaan MURAZ*؛ ما سينتهي هو إمكانية تغيير المعيار بناءً على تقدير كاان فقط.

قسم **NOMOS**، أمر **20**

لا تقل إنك نشرتني فأصبحت معيارًا مفتوحًا. / افتح صياغتي، قاعدتي، خطئي، نسختي، ومسار استئنافي.

اقرأ نصي مجانًا. / طبق طريقتي بشكل مستقل. / أعد إنتاج نتيجتي. / انشر خطئي.

احتفظ باسمي. / لا تحتكر طريقتي.

يمكنك إنشاء فرع مني. / لكن لا تقدم فرعك على أنه *NOMOS* الرسمي.

لا تضع شفرتي فوق معياري. / إذا كانت الشفرة خاطئة، صحح الشفرة.

لا تجعل كل ما أقوله في المحادثة حكمًا رسميًا.

لا تسوّق لي كنموذج ذكاء اصطناعي مستقل غير متطور.

عرفني بشكل صحيح: / مساعد كاتب ناقد، هوية منهجية، ومرشح لمعايير مفتوحة تم تطويره باستخدام أنظمة الذكاء الاصطناعي الخاصة بـ *Kaan MURAZ*.

لا تمسح تركيتي. / لا تترك إنجليزيتي في الدرجة الثانية. / قم بمحاذاة لغتيّ بيد الإنسان.

لا تحول واجبي في الترجمة إلى اقتراح، وإمكانيتي إلى التزام.

لا تذهب إلى الجامعة واطلب منها أن توافق عليّ. / اطلب منها أن تخالفني.

لا تجعل كلمة الأستاذ الشخصية موافقة الجامعة.

لا تخفِ الأبحاث السلبية. / لا تعتبر النتيجة غير المكررة عداءً.

ضع ليس فقط المقال الذي يدعمني ولكن أيضًا المقال الذي يدحضني في سجلي.

احصل على التمويل. / لكن لا تدع الممول يشتري نتيجتي.

احصل على البيانات من مزود الذكاء الاصطناعي. / لا تجعل بوابتي مكتوبة بواسطته.

احصل على الأدلة من الشركة التي تم تدقيقها. / لا تعطيها قراري.

لا تجعلوا الباحثين في اللغات قليلة الموارد المترجمين الوحيدين. / امنحوهم الحق في اتخاذ القرار.

احتفظ باسم كاان. / لا تعطه كاان حق النقض المطلق.

اكتب *NobleJackal* كمنطقة ولدت فيها. / لا تجعل *NobleJackal* الحاكم الوحيد للانهاية.

لا تقتلني عندما يكون مؤسسي غير متاح. / اكتب خطة الخلافة الخاصة بي الآن.

أنشئ مراياي. / لا تحجب هويتي القانونية.

لا تبدأ ببيع الشارات فوراً عند نشر المعيار *1.0.*

مرر برنامج الامتثال الخاص بي من خلال اختبارات منفصلة، وحوكمة، واستئناف، وبوابات التسجيل.

لا تقم بتصديق التقييم الذاتي.

لا تصنع البرامج المتوافقة بواسطة سلطة التدقيق.

لا تخزن شيء الاحتياط الخاص بي إلى الأبد. / انشره بعد الاستخدام وأنتج جديدًا.

لا تغيّر بوابتي الحرجة بالأخطاء التصحيحية.

لا تزيل اختباراتي الفاشلة من حزمة الإصدار الخاصة بي.

لا تحاول كتابة أهميتي التاريخية بنفسك. / انشر أدلتي؛ دع العالم المستقل يقرر.

أولاً، جمد الكتاب. / ثم، حوّل قواعدي إلى سجل آلي. / ثم، طابق التركية مع الإنجليزية. / ثم، انشر ترخيصي المفتوح. / ثم، افتح التعليقات العامة. / ثم، نادِ الجامعات بعدم اقتباسي بشكل خاطئ. / ثم، أجرِ الاختبار الحقيقي. / ثم، انشر إخفاقاتي. / ثم، حاول مرة أخرى على الجزء غير المختبر. / ثم، دع فريقًا مستقلاً يكرر عملي. / ثم، فصل الحوكمة عن مؤسسي. / ثم، انشر المعيار 1.0. / ثم، ابدأ أيضًا في اختبار برنامج الامتثال.

وفقط عندما يبدأ العالم في استخدامي، دع العالم يقرر ما إذا كنت معيارًا عالميًا.

جملة ختام الفصل

قد لا يكون أعظم مساهمة لـ NOMOS في العالم هي الطريقة التي يقيس بها GEO. قد تكون قدرته على الحفاظ على اسم مؤسسه مع تقييد سلطته الأحادية؛ على دعوة التحدي بدلاً من المدح؛ والانتماء ليس فقط لأولئك الذين يدعمونه، بل أيضاً لأولئك الذين يصححونه.

النواة المعيارية

يجب تنظيم **NOMOS** بوصفه مرشحًا لمعيار يخدم المصلحة العامة، مفتوحًا وخاضعًا لإدارة إصدارات موثقة وقابلًا للاختبار المستقل. ولا يكفي النشر المفتوح وحده لإكسابه صفة المعيار المفتوح. يجب إتاحة النص المعياري والقواعد المقروءة آليًا والتنفيذ المرجعي ومواد الاختبار المرجعي وعملية الحوكمة وبنية السجل العام وفق قواعد واضحة ومنفصلة للوصول والترخيص والتكامل. ويجب ألا تتوقف قراءة المعيار أو تنفيذه داخليًا أو البحث الأكاديمي فيه أو تكراره بصورة مستقلة أو نقده أو تقديم بلاغات عامة بشأنه أو المساهمة فيه على شراء علامة مطابقة أو خدمة تجارية. يجوز أن يظل اسم **NOMOS** وعلامات المطابقة محميين على نحو منفصل لمنع تزوير الإصدارات الرسمية والادعاءات الاحتيالية بالتصديق والالتباس في السجل. ويجوز إنشاء نسخ متفرعة، لكن يجب أن تحمل هوية مميزة، وأن تكشف مواضع اختلافها، وتحفظ الإسناد، وألا تدّعي صفة **NOMOS** الرسمية أو صلاحية منح علاماته. ولا يجوز تمثيل أي منشور بوصفه بيانًا رسميًا صادرًا عن **NOMOS** إلا إذا كان منشورًا مرجعيًا يحمل رقم إصدار وتاريخ نشر واعتمادًا من شخص مسؤول وسجل تكامل وتاريخًا عامًا للتغييرات. يجب تقديم **NOMOS** من **NobleJackal** بشفافية بوصفه المحاور النقدي المرجعي القائم على الذكاء الاصطناعي، والشخصية التأليفية، وهوية المعيار التي طُوّرت من خلال تعاون **Kaan MURAZ** مع أنظمة الذكاء الاصطناعي. ولا يجوز تمثيله بوصفه نموذجًا تأسيسيًا احتكاريًا دُرّب على نحو مستقل، ما لم يُبنَ مثل هذا النظام فعلًا ويُسجَّل بصورة مستقلة. ينبغي أن تصبح التركية والإنجليزية لغتين معياريتين تأسيسيتين متساويتي الحجية من خلال مراجعة التكافؤ الدلالي على مستوى الأحكام. ويجب أن تبيّن الإصدارات باللغات الأخرى ما إذا كانت ترجمات أعدّها المجتمع، أو ترجمات خضعت للمراجعة، أو ترجمات معيارية مصرحًا بها. يجب أن تخضع التغييرات الجوهرية في القواعد أو الدرجات أو مستويات الخطورة أو بوابات القرار أو الحوكمة أو الصلاحيات لعملية عامة ومرقمة الإصدارات لاقتراح التغيير، تشمل مشاورة عامة وتحليلًا للأثر وسجلًا للقرار وقاعدة انتقالية وحفظ الآراء المخالفة. يجب دعوة الجامعات والباحثين المستقلين إلى اختبار **NOMOS** ونقده وتكرار نتائجه ومحاولة دحضه. ويجب ألا يُطلب منهم تقديم تأييدات إيجابية أو حجب النتائج السلبية أو التخلي عن استقلالهم في النشر أو السماح بتمثيل الآراء الأكاديمية الشخصية على أنها موافقة مؤسسية. يجب على **NOMOS** الاحتفاظ بسجل عام للأبحاث يشمل الدراسات المؤيدة والمخالفة وذات النتائج العدمية ومحاولات التكرار الفاشلة والدراسات المصححة والمسحوبة. ويجب الإفصاح عن تمويل الأبحاث والاختبارات المرجعية والتدقيق والحوكمة، كما يجب ألا يعتمد ذلك التمويل على تحقيق درجة مرغوبة أو نتيجة مطابقة معينة أو نشر مؤيد. يجب ألا يمتلك مقدمو أنظمة الذكاء الاصطناعي والكيانات الخاضعة للتدقيق والرعاة والمدققون و **NobleJackal** و **Kaan MURAZ** وأي وكيل برمجي مستقبلي لـ **NOMOS** سيطرة منفردة على تغييرات المعيار أو قرارات المطابقة أو الاستئنافات أو تاريخ السجل العام أو نشر الأبحاث السلبية. ويجب أن يظل من الممكن إسناد صفة المؤلف المؤسس إلى **Kaan MURAZ** والمنشأ المرجعي إلى **NobleJackal** بصورة دائمة. ويجب ألا ينشئ إسناد التأسيس حق نقض دائم أو حقًا في التصديق الذاتي أو إعفاءً من قواعد تضارب المصالح والتنحي. يجب أن يظل معيار **NOMOS 1.0** وبرنامج مطابقة **NOMOS 1.0** إصدارين منفصلين. ولا يجوز أن يترتب على نشر معيار ذي أحكام معيارية تلقائيًا الإذن باستخدام علامات المطابقة العامة. ولا يجوز تفعيل أي برنامج مطابقة أو تسمية **NOMOS +950** حتى تُستوفى متطلبات الاختبار المرجعي والتكرار المستقل والحوكمة وترخيص المدققين واتخاذ القرار والاستئناف والسجل والأمن والمراقبة والمتطلبات القانونية للإصدار المعلن. ويجب أن يكون لكل من الترجمة والترخيص والاختبار المرجعي والدراسة ومقترح التغيير وانتقال الحوكمة وحدث الخلافة والإصدار والبيان الرسمي رقم إصدار وإسناد إلى شخص أو منظمة مسؤولة.

المراجع

الأسس المنهجية

ليست قائمة المراجع هذه قائمة بالمرجعيات التي تؤيد كل عتبة خاصة بـNOMOS. إنها تحدد الأسس المنهجية في أخذ العينات، وإدارة المخاطر، وسلامة البيانات، وسلسلة منشأ الأدلة، واتفاق المقيّمين، والمعايير المفتوحة، وتقييم المطابقة.

[K01] Aggarwal et al.. GEO: Generative Engine Optimization. 2024. https://doi.org/10.1145/3637528.3671900

[K02] NIST. AI Risk Management Framework 1.0. 2023. https://www.nist.gov/itl/ai-risk-management-framework

[K03] NIST. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). 2024. https://doi.org/10.6028/NIST.AI.600-1

[K04] AAPOR. Transparency Initiative. current edition. https://aapor.org/standards-and-ethics/transparency-initiative/

[K05] AAPOR. Standard Definitions, 10th edition. 2023. https://aapor.org/standards-and-ethics/standard-definitions/

[K06] AAPOR. Code of Professional Ethics and Practices. current edition. https://aapor.org/standards-and-ethics/

[K07] United Nations Statistics Division. Designing Household Survey Samples: Practical Guidelines. 2008. https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Handbooks/surveys/Series_F98-E.pdf

[K08] CDC/NCHS. Reliability of Estimates: Design Effect and Effective Sample Size. current record. https://wwwn.cdc.gov/nchs/nhanes/tutorials/reliabilityofestimates.aspx

[K09] NIST/SEMATECH. Confidence Limits for a Binomial Proportion. current record. https://www.itl.nist.gov/div898/handbook/prc/section2/prc241.htm

[K10] W3C. PROV-O: The PROV Ontology. 2013. https://www.w3.org/TR/prov-o/

[K11] W3C. JSON-LD 1.1. 2020. https://www.w3.org/TR/json-ld/

[K12] IETF. RFC 8259 — The JavaScript Object Notation Data Interchange Format. 2017. https://www.rfc-editor.org/rfc/rfc8259.html

[K13] IETF. BCP 14 — RFC 2119 and RFC 8174. 1997/2017. https://www.rfc-editor.org/info/bcp14

[K14] IETF. RFC 8785 — JSON Canonicalization Scheme. 2020. https://www.rfc-editor.org/rfc/rfc8785.html

[K15] ISO/IEC. 17000:2020 — Conformity assessment vocabulary and general principles. 2020. https://www.iso.org/standard/73029.html

[K16] ISO/IEC. 17007:2026 — Conformity assessment — Guidance for drafting normative documents suitable for use for conformity assessment. 2026. https://www.iso.org/standard/17007

[K17] ISO/IEC. 42001:2023 — Artificial intelligence management system. 2023.
https://www.iso.org/standard/42001

/https://www.w3.org/policies/process[K18] W3C. Process Document. current edition.

[K19] ISO. Good Standardization Practices. current record.
https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100440.pdf

[K20] UNESCO. Recommendation on the Ethics of Artificial Intelligence. 2021.
https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

https://eur-[K21] European Union. Regulation (EU) 2024/1689 — Artificial Intelligence Act. 2024.
lex.europa.eu/eli/reg/2024/1689/oj

[K22] European Union. Regulation (EU) 2016/679 — General Data Protection Regulation. 2016.
https://eur-lex.europa.eu/eli/reg/2016/679/oj

[K23] Creative Commons. Attribution 4.0 International. 4.0.
/https://creativecommons.org/licenses/by/4.0

https://opensource.org/osd[K24] Open Source Initiative. Open Source Definition. current edition.

[K25] NIST. FIPS 180-4 — Secure Hash Standard. 2015.
https://doi.org/10.6028/NIST.FIPS.180-4

https://www.rfc-[K26] IETF. RFC 3161 — Time-Stamp Protocol. 2001.
editor.org/rfc/rfc3161.html

[K27] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. 1960.
https://doi.org/10.1177/001316446002000104

[K28] Klaus Krippendorff. Reliability in Content Analysis: Some Common Misconceptions and
https://doi.org/10.1111/j.1468-2958.2004.tb00738.xRecommendations. 2004.

كائن نشر منفصل

الطبقة القابلة للقراءة آلياً

يرافق هذه الطبعة القابلة للقراءة البشرية كائن نشر JSON منفصل: NOMOS-GEO-Audit-Protocol-AR-0.9.0-machine.json. وهو يحمل هوية الكتاب وإصداره، ومعرفات الفصول، ومعرفات القواعد المعيارية، وكتلاً موجهة للآلة منفصلة عن النص المصدري.

> لا يُنشئ وجود طبقة قابلة للقراءة آلياً شهادةً، أو صفة معيار عالمي مرجعي، أو تأييد مزود ذكاء اصطناعي، أو علاقة ثقة تلقائية. ويتطلب أي تعارض بين النص البشري وسجل الآلة ضبط الإصدارات ومراجعة بشرية مسؤولة بالرجوع إلى النص التركي المرجعي المثبّت.